View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Vanderbilt Electronic Thesis and Dissertation Archive

A COMPARISON OF DYNAMIC ASSESSMENT AND PROGRESS MONITORING IN THE PREDICTION OF READING ACHIEVEMENT FOR STUDENTS IN KINDERGARTEN AND FIRST GRADE

By

Erin Caffrey

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Special Education

December, 2006

Nashville, Tennessee

Approved:

Professor Douglas Fuchs

Professor Lynn S. Fuchs

Professor Donald L. Compton

Professor Daniel H. Ashmead

For Paxton.

ACKNOWLEDGEMENTS

I would like to acknowledge and thank several deserving people who mentored me through the dissertation process. I would like to thank my committee members for their time, patience, expertise, and counsel. First, I would like to express my gratitude to my advisor, Doug Fuchs, who has offered me his time and wisdom. I am continually impressed by his magnitude of vision and dedication to his work. To Lynn Fuchs, whose efficiency and work ethic are unparalleled. To Don Compton, who always makes himself available to students despite his busy schedule. And to Dan Ashmead, who helped me create a foundation of knowledge that guided me through my program and will continue to guide me through my career.

I would also like to acknowledge the work of two wonderful project coordinators. To Loulee Yen, who has given me counsel, understanding, and friendship for the last four years. And to Bobette Bouton, who always gives more than necessary. I will always admire their energy and organization.

Last, but certainly not least, I would like to thank Andrea and Natalie, who supported me through numerous joys and disappointments and helped me maintain a healthy sense of reality.

TABLE OF CONTENTS

Appendix

LIST OF TABLES

CHAPTER 1

NATURE OF THE PROBLEM

Assessment in education is a process of collecting data for the purposes of making decisions. Data from traditional assessment reflects students' learning because it is typically administered by a neutral examiner who does not give performance-contingent feedback. Traditional assessment offers no scaffolding or social support for learning. Data from dynamic assessment (DA), by contrast, represents both the process and product of students' learning. DA is administered by an examiner who provides scaffolding, social support for learning, and intervention when a student fails. In other words, whereas traditional assessment measures independent performance (i.e., product), DA measures both independent performance and assisted performance (i.e., process). Independent performance is what the student can achieve alone; assisted performance is what the student can achieve with the help of the examiner.

In the identification of students at-risk for reading failure, DA may provide useful information. It controls for unequal background knowledge when assessing young students who are entering school with varied experiences. As discussed, DA measures (a) students' independent performance, which represents the accumulated knowledge of their experiences, and (b) their assisted performance, which represents their potential achievement if given adequate instruction. Presumably, students' assisted performance is an indication of their ease of learning, or how well they will achieve during standard classroom instruction. If students have a low level of assisted performance, they may require more intensive instruction than the general education classroom can provide.

Theoretical Origins of DA

DA is grounded in the theory of Vygotsky's social constructivism. In social constructivism, a child's learning occurs through participation in socially or culturally embedded experiences with a more experienced adult. According to Vygotsky, learning takes place within the *zone of proximal development* (ZPD). The ZPD is the range of learning children can achieve while engaged in meaningful activities with a more experienced adult. It is measured as the difference between what the child can accomplish alone and what the child can accomplish with *scaffolding.* Scaffolding is a support system set up by the adult to guide the child through the learning process. For example, an adult may use the child's knowledge of addition to teach principles of multiplication. The adult provides more and more support until the child is able to connect "new" information to "known" information.

The ZPD has pedagogical implications in the classroom. Some believe that good teaching should include helping a student through the ZPD (Tharp & Gallimore, 1988). Teaching, therefore, is considered a constant negotiation between student and teacher. As students learn, they become more responsible for their learning and self-regulation. If students fail to self-regulate their learning despite scaffolding, the teacher must go back to instruction at a lower level of cognitive development.

Types of DA

DA methods differ in terms of their purposes and procedures. The most commonly used DA methods include Feuerstein's Learning Potential Assessment Device (LPAD), Budoff's Learning Potential Testing, graduated prompts, the information-processing framework, and testing-the-limits procedures. Each of these methods is briefly described, highlighting their salient features.

Feuerstein's LPAD

Feuerstein's LPAD is a process of mediated learning that focuses on changing deficient cognitive processes in students who have difficulty learning. The LPAD was designed to develop a child's *cognitive modifiability* – an independent ability to self-modify cognitive processes and adapt to changing demands (Grigorenko & Sternberg, 1998). Examiners are trained to alter the administration of test items in four ways: the structure of the instruments, the nature of the test situation, the orientation to process, and the interpretation of results (Feuerstein, Rand, & Rynders, 1988). The LPAD consists of both verbal and nonverbal subtests that focus on skills such as reasoning, categorization, and memory strategies. Although the LPAD is an assessment device, Feuerstein's primary purpose is remediation.

Extensive research using the LPAD has been conducted. Much of it, however, did not use control groups and has not been published in peer-reviewed journals. Very generally, researchers found that (a) performance on LPAD posttest is higher than LPAD pretest, (b) a longer mediation period leads to greater gains, and (c) disadvantaged students benefit more from the LPAD than advantaged students (Grigorenko & Sternberg, 1998). Use of the LPAD in research has two distinct disadvantages. First, it is time consuming. The initial administration of the LPAD can take up to 10 hours for each participant. Second, reliability and validity of the instrument have not been explored extensively by Feuerstein. It was only after Feuerstein that researchers made an effort to standardize protocols and explore issues of reliability and validity.

Budoff's Learning Potential Testing

Budoff's learning potential testing is also known as test-train-test assessment (Grigorenko & Sternberg, 1998). *Learning potential tests* were designed as an intelligence measure, specifically for disadvantaged students. The assessment is a standardized coaching technique that redirects students' attention to a problem, explains crucial attributes of a problem, and offers continuous praise and encouragement. Coaching continues until mastery is reached. Budoff's measure of learning potential is unique in that it was designed specifically for disadvantaged students and for the purpose of educational placement.

Budoff and his colleagues conducted numerous studies concerning the validity of his instrument. To provide evidence of construct validity, Budoff and colleagues reported in many studies that coaching leads to improvement on posttest (Budoff, 1967, 1987a, 1987b, Budoff & Friedman, 1964). In terms of predictive validity, Budoff demonstrated that learning potential predicts both teacher ratings of achievement (Budoff, 1987a, 1987b) and classroom performance (Budoff, Meskin, & Harrison, 1971). In fact, learning potential was found to be the best predictor of classroom achievement for children enrolled in special education programs (Budoff, Corman, & Gimon, 1976; Budoff et al., 1971).

Graduated Prompts

The graduated prompts method was developed by Campione and Brown (Campione & Brown, 1987). It is alternatively referred to as *testing through learning and transfer*. The graduated prompts method sets up a system of scaffolding in which the students are given a series of progressively explicit hints until they can solve a problem independently. The hints are standardized and administered in a predetermined order. *Learning ease* is operationalized as the number of hints necessary for success on a problem. Students who require the fewest number of hints are believed to have the greatest learning ease. In addition, maintenance and transfer are often measured to assess a student's ability to use learned information flexibly or in new contexts (Campione, Brown, & Bryant, 1985).

A primary interest of Campione and Brown was investigating cognitive differences between students of low ability compared to those of high ability. Results indicate that students of low ability require more hints to solve a problem and transfer information than students of high ability (Campione, Brown, Ferrara, Jones, & Steinberg, 1985; Ferrara, Brown, & Campione, 1986), evidence of construct validity. A secondary interest was determining the extent to which a graduated prompts DA measure could predict future achievement. Campione and Brown (1987) used a matrix reasoning task and a series completion task to determine the amount of achievement gain variance accounted for by DA training, DA transfer, and IQ. DA training and transfer score on both matrix reasoning tasks and series completion tasks contributed significantly to the variance in achievement gain. IQ was found to be a significant, yet inconsistent, predictor of achievement gain. Although there is evidence suggesting that graduated prompts DA can predict future achievement on a posttest measure, it is unknown if graduated prompts DA can predict future academic achievement.

Information-Processing Framework

The majority of DA research using the information-processing framework was conducted by Swanson. He developed the Swanson Cognitive Processing Test (S-CPT), which is a standardized dynamic instrument that measures processing abilities. The primary process thought to contribute to learning is working memory. Therefore, children's difficulties in skill acquisition and learning are attributed to deficits in working memory. The S-CPT measures *processing potential*, which is analogous to Feuerstein's concept of cognitive modifiability. Processing potential is operationalized through the measurement of seven scores: initial score, gain score, probe score, maintenance score, processing difference score, processing stability score, and strategy efficiency score.

The S-CPT is designed to investigate two questions: (1) Do children with learning disabilities have generalized or specific working memory deficits compared to average achieving children? and (2) What is the degree of modifiability of working memory performance in children with learning disabilities? Swanson has reported that various S-CPT scores are significant predictors of achievement and classification, however, results are inconsistent. In addition, Grigorenko & Sternberg (1998) have questioned the interpretation of his data.

Testing-the-Limits Procedures

Carlson and Wiedl (1978, 1979) developed testing-the-limits procedures by combining their empirical findings with information-processing theory. They believe that test performance is a combination of the individual student, the test materials, and the test situation. The testing-the-limits approach focuses on the test situation. Examiners use conventional assessment measures, but they are trained to manipulate the test environment to improve the performance of students with learning problems.

Empirical research conducted by Carlson and Wiedl (and others) has focused on what kinds of testing conditions elicit optimal performance for different types of students. Students are grouped in pre-determined categories and taught as a group. In general, verbalization and elaborated feedback testing conditions were the most effective, especially for students with low ability, students with high anxiety, or on difficult test items that require high levels of cognitive processing. Due to the group administration of the testing-the-limits procedure, individual comparison is not possible. Therefore, the validity of the results depends heavily on the validity of the pre-determined categories (i.e., high anxiety vs. low anxiety).

DA in Today's Research Environment

With the current trend of empiricism in educational research, certain types of DA are more compatible than others with respect to today's standards of rigor. Current educational standards in research and practice seem to value standardization of protocols, reliability of measurement, fidelity of testing procedures, efficiency, and utility on a broad scale. DA methods that are more treatment-oriented, such as Feuerstein's LPAD, are often lengthy, highly individualized, and may not generalize to a broader population. In addition, developers of DA that is a treatment have been less concerned with standardization, reliability, and fidelity. Approaches to DA as a treatment have been designed to benefit the child directly and elicit immediate change in the child's cognitive or educational functioning.

Alternatively, more assessment-oriented DA methods, such as graduated prompts, are often efficient, standardized, and have the potential to generalize to a broader population. Such DA methods are not necessarily designed to provide a direct benefit to the child during the testing session. Instead, DA is viewed as a tool to estimate current ability, predict future academic ability, or design interventions.

To further the use of either DA method (i.e., for treatment or assessment), research must first be conducted to validate the DA instrument itself. Without a valid assessment device, treatment tends to be unfocused and may be misguided. A good starting point to validate a DA instrument would be to use an assessment-oriented instrument with adequate measurement properties. This instrument would permit quantitative analyses to explore reliability and validity. For the current study, graduated prompt DA was selected to investigate issues of reliability and validity.

Dissertation Purpose and Research Questions

The purpose of the proposed research is to study the predictive validity of DA in comparison to that of two other common screening methods: initial performance measures and progress monitoring. The predictive validity of various screening measures has important implications for educational practice. Because of dissatisfaction with the use of IQ tests and discrepancy models over the past 10 to 15 years, researchers and educators have been investigating more efficient methods of early identification of students who are at-risk for school failure. If DA, initial performance measures, or progress monitoring are found to have predictive validity, they have the potential to substantially reduce the time teachers need to identify at-risk children. Furthermore, these three screening methods may be able to lower at-risk students' exposure to repeated school failure. The relative predictive validity of DA, initial performance, and progress monitoring measures, however, is unknown. It is important to understand the relative utility of the three screening measures so that educators can use them appropriately.

The research questions guiding the study are as follows:

1. *Is Fall DA score a significant predictor of Spring reading achievement? Is initial performance on single word reading measures a significant predictor of Spring reading achievement? Is progress monitoring over a five week period a significant predictor of Spring reading achievement? Which is strongest?*
2. *Do Fall DA score, initial performance on single word reading measures, and progress monitoring over a five week period explain unique variance in Spring reading achievement?*

CHAPTER II

REVIEW OF LITERATURE

The purposes of educational assessment are to evaluate current achievement, predict future achievement, and prescribe educational treatments. Conventional one-point-in-time assessment (i.e., “static”) or traditional pretest-posttest assessments have been used to accomplish these aims because they are standardized, easily administered, and norm-referenced. Traditional assessment produces clear-cut results that are used to evaluate, identify, and classify children. Nevertheless, many believe it should not be used for these “high-stakes” purposes. Traditional assessment has been criticized for underestimating general ability (Swanson, 2001) and lacking sensitivity toward so-called disadvantaged students (e.g., Peña, Quinn, & Iglesias, 1992; Utley, Haywood, & Masters, 1992) and students with disabilities (e.g., Lidz, 1987). Ironically, traditional assessment is often used to identify and place low-achieving, at-risk students. Scores on traditional assessment tests are difficult to interpret for low-achieving students because of floor effects. Many students in kindergarten or first grade are unskilled readers. When given a traditional reading assessment, such as the WRMT-R Word ID and Word Attack, a high proportion of these students will receive a score of zero. How should a score of zero be interpreted? Is a score of zero indicative of an unskilled reader who is not yet ready to acquire those skills easily? Or, is a score of zero indicative of a currently unskilled reader who is ready to become skilled? Dynamic assessment (DA) is a possible alternative to traditional assessment that can begin to tease apart these two groups of unskilled readers.

Alternative to Static Assessment

DA has been defined and operationalized in different ways, such as learning potential assessment (e.g., Budoff et al., 1971, 1974); mediated learning experience (e.g., Feuerstein et al., 1979); testing-the-limits procedures (Carlson & Wiedl, 1978, 1979); mediated assessment (e.g., Bransford, Delclos, Vye, Burns, & Hasselbring, 1987); and graduated prompts (e.g., Campione, Brown, Ferrara, Jones, & Steinberg, 1985). DA differs from traditional assessment in terms of the nature of the examiner/student relationship, the content of the feedback, and the emphasis on process, rather than product (Grigorenko & Sternberg, 1998).

DA vs. Traditional Assessment: An Overview of Differences

In traditional assessment, the examiner is a neutral or "objective" participant who provides only standardized directions. In DA the examiner attempts to form a closer relationship with the student that will foster learning. In traditional assessment, the examiner does not give performance-contingent feedback. Indeed, the traditional assessment examiner is often explicitly discouraged from making any statements that may alter the independent achievement of the student. In DA, the examiner not only gives performance-contingent feedback, but offers instruction in response to student failure to alter or enhance student achievement. In short, traditional assessment is oriented towards the product of student learning (or performance), whereas in DA the interest is in both the product and process of student learning (or rate of growth).

The different DAs may differ more among themselves than any one particular variant differs from traditional assessment. The DA procedures each have their own theoretical bases, purposes, and procedures. Some, like learning potential assessment, mediated learning, and mediated assessment, are characterized by a strong clinical orientation with an emphasis on instruction. Others, like testing-the-limits procedures and graduated prompts, can claim a strong research orientation and an emphasis on achievement prediction, educational placement, and prescription of intervention techniques. This classification of the various DA procedures does not preclude research on "clinically-oriented" DA nor the use of "research-oriented" DA in clinical practice. Because of the variety of DA procedures, it is difficult if not impossible to offer a single, all-encompassing definition. In general, DA investigates the change in student performance brought about by deliberate examiner intervention. The performance change due to this examiner intervention is used as a presumably unbiased measure of current achievement, to predict future achievement, and to inform intervention.

Proponents of DA claim it is based on the often ignored link between assessment and intervention by measuring both the process and product of student learning. For example, some students may enter kindergarten with little background knowledge. These students may score low on traditional assessment. But if they possess the intelligence, behavioral maturity, and motivation necessary for learning, they may score higher on DA. Such a child may be in less danger of school failure than one who scores low on both traditional assessment and DA. The pattern of low traditional assessment score *and* low DA score may truly represent those students who are most likely to experience school failure. In addition to their predictive information, prescriptive data can be derived to identify the type and intensity of intervention that is required for success. DA incorporates a test-teach-test format, conceptually similar to response-to-intervention (RTI) techniques. However, DA can potentially measure RTI within a much shorter time frame.

Clinically-Oriented vs. Research-Oriented DA

The broad practice of DA has evolved and diverged into two separate strands of study: clinically-oriented DA and research-oriented DA. Clinically-oriented DA began as an educational treatment or intervention. Its most common operationalization is Feuerstein's Learning Potential Assessment Device (LPAD). The LPAD is a nonstandardized method of assessing and treating cognitive deficiencies in children with learning problems. Treatment duration could last years (Rand, Tannenbaum, & Feuerstein, 1979). Research-oriented DA, by contrast, originated as an instrumental tool. It is generally (although not always) a standardized administration of an assessment in which the examiner seeks to guide the student through the learning process during one teaching session. The time needed for a student to reach mastery, or the necessary level of instructional explicitness, serves as an index of student potential.

Three concerns about DA are typically expressed: It is weakened by construct fuzziness; research has only infrequently explored its technical characteristics; and it is labor intensive. These criticisms are discussed separately for clinically-oriented and research-oriented DA.

Construct fuzziness. Construct fuzziness (Jitendra & Kame'enui, 1993) refers to when DA's theoretical bases, purposes, procedures, and uses do not have a unified focus. "Fuzziness" often occurs when researchers fail to make the distinction between clinically-oriented or research-oriented. The purpose of clinically-oriented DA is to remediate the deficient cognitive processes that contribute to learning problems. Procedures are generally not standardized and require the examiner to function as an educator. Moreover, the educator must rely heavily on insight and expertise to assess learning problems and adapt intervention. This type of DA is used to improve student achievement directly. The purpose of research-oriented DA is early identification and classification of students with learning problems. DA procedures are often standardized and relatively easily implemented by trained examiners. Research-oriented DA, by contrast, may or may not directly improve student achievement. It is used as a tool to identify those students who require more intensive intervention and to place them in a setting in which that intervention can occur.

However, research on clinically-oriented and research-oriented DA has not thrived due partly because there is no agreement on what constitutes these two strands of DA in the literature. Consequently, research on either strand is difficult to identify, synthesize, and extend.

Technical characteristics. Research in extant literature does not report reliability and validity data on the specific DA measures used. In addition, many types of DA are not standardized, and fidelity of implementation is not reported, leaving readers to question the accuracy and consistency of its implementation. Part of this problem stems from the lack of standardization in many DA procedures. Without standardized procedures, technical characteristics cannot easily be studied. The debate over standardization is a good example of the trade-off between clinically-oriented DA and research-oriented DA. Proponents of clinically-oriented DA believe standardization contradicts its spirit and theoretical orientation (e.g., Feuerstein, 1979). That is, a standardized approach would fail to provide truly individualized intervention in response to student failure. Proponents of research-oriented DA believe standardization and technical adequacy are necessary to make it a worthwhile tool for research and practice (e.g., Swanson, 1994; Bryant, Brown, and Campione, 1983; and Ferrara, 1987). Due to lack of standardization, the technical characteristics of clinically-oriented DA are very difficult to study. And, although research-oriented DA protocols are more likely to be standardized, the technical characteristics have not been studied adequately.

Labor intensive. Some critics have suggested that the time required to develop new test protocols and train examiners may not be worth the information DA provides. Traditional standardized assessments have already been developed over a number of years and training examiners is straightforward. DA protocols have been in use for decades, too, but because of their lack of technical adequacy, more time may be needed to establish the validity standards expected in today's educational research.

Again, this criticism may be moderated by the type of DA orientation. Clinically-oriented DA requires relatively little time to develop test protocols because scripted protocols are not essential. Only a general framework of scaffolding serves as a protocol. Because of this, educator insight and expertise are essential to the successful implementation of DA. As the intervention becomes less standard, student achievement becomes more dependent on the specific educator who provides the intervention. Educators who provide clinically-oriented DA must be proficient in many types of intervention and have the ability for "on-line" problem solving in order to switch types of intervention when the student fails to respond. Conversely, research-oriented DA requires an extensive amount of time to develop test protocols because they must be standardized and possibly normed based on a sample of the target population. The demand for educator insight and expertise, however, is much lower than in clinically-oriented DA. Because procedures are standardized, educators can be trained in a time-frame similar to that of traditional assessment.

Is There a Need for DA?

Currently, DA is not a viable alternative to traditional assessment. Some believe DA should not replace traditional assessment, but rather be used in conjunction with it (e.g., Lidz, 1987). The question then becomes, "What unique information can DA provide?" First, DA may offer a less-biased measure of achievement for certain populations because it is less dependent on mainstream language skills and background experience (e.g., Sewell, 1979; Sewell & Severson, 1974; Peña et al., 1992). It may be especially useful to differentiate various low-achieving students. As discussed, traditional tests are often subject to floor effects for low-achieving students. Items are scored "right" or "wrong" using an all-or-nothing mentality. DA, by contrast, gives multiple opportunities for success. Low-achieving students, therefore, can be differentiated along the continuum of how easily they learn.

Second, clinically-oriented DA may inform instruction so that educational interventions can be more readily designed (e.g., Feuerstein, 1979; Haywood, 1992). If a test is susceptible to floor effects and students fail all items, we do not have useful data to gauge their academic functioning and plan appropriate interventions. And third, research-oriented DA has the potential to predict future student achievement because it attempts to measure the process and ease of learning. Presumably, those who learn with more ease will benefit more from classroom intervention and achieve at a higher level. Research-oriented DA can be used to predict achievement within the context of an RTI framework. Students' responses to teaching during DA may approximate how they will respond to longer-term classroom intervention. DA has the potential to offer a faster alternative to RTI identification procedures.

Purpose of Review

This review focuses on the ability of DA to predict future achievement. Several extensive reviews of DA are available in the extant literature (e.g., Grigorenko & Sternberg, 1998; Swanson, 2001). Grigorenko and Sternberg (1998) offer a comprehensive descriptive review that examines types of DA, broadly defined, based on their comparative informativeness, power of prediction, degree of efficiency, and robustness of results. Although the review is comprehensive, no quantitative syntheses were conducted and DA's predictive validity was not systematically analyzed. Swanson (2001) conducted a selective quantitative synthesis of DA. He used effect sizes (ESs) and mixed regression analyses to model responsiveness to DA, and found that the magnitude of the ESs was best predicted by type of DA and assessment domain. In general, his analysis focused on differences between ability groups and effectiveness of different types of DAs as assessments or interventions. He did not pursue issues of validity.

Prediction of future achievement is important because it may identify the students who are at-risk for school failure and need more intensive intervention. Students enter school with different abilities based on differences in intelligence, home experiences, and prior education. These abilities and experiences result in different levels of academic competence upon entering kindergarten. At this time, traditional assessment will reflect mostly a student's current knowledge but not learning potential. In this scenario, DA that indicates a student's potential for change when receiving instruction may be used in conjunction with traditional assessment to determine the likeliness of school failure and plan appropriate instruction.

Method

Definitions

As indicated, no single definition of DA exists. In this review, *dynamic assessment* refers to any procedure that examines the effects of deliberate, short-term, intervention-induced changes on student achievement, with the intention of measuring both the learning process and product. In addition, the DA must provide corrective feedback and intervention in response to student failure. As discussed, DA is used for many purposes: to measure current achievement, to predict future achievement, and to inform intervention. This synthesis is concerned primarily with the predictive validity of DA; that is, how well does DA predict future student achievement?

Inclusion Criteria

Four inclusion criteria were used to select articles for this review. First, included articles were published in English. Several relevant lines of research in DA have been published in Russian (e.g., Ginzburg, 1981; Goncharova, 1990; Vlasova, 1971), German (e.g., Carlson & Wiedl, 1980; Guthke, 1977; Wiedl & Herrig, 1978), and Dutch (e.g., Hamers, Hessels, & Van Luit, 1991; Hamers & Ruijssenaars, 1984). A subset of these authors published a collection of studies in English which were included in this review

(Hessels & Hamers, 1993; Meijer, 1993; Reising, 1993; Tissink, Hamers, & Van Luit, 1993). If only secondary reports were available in English, these studies were excluded (e.g., Flammer & Schmid, 1982; Hamers & Ruijssenaars, 1984).

Second, articles included participants between preschool and high school. A study by Shochet (1996), for example, was excluded for using South African college students. Third, articles included students with high-incidence disabilities, students at-risk for school failure due to cultural or economic disadvantage, second language learners, or normally achieving students. Students with low-incidence disabilities, such as sensory impairments, were not included in this review (e.g., Dillon, 1979; Tellegen & Laros, 1993).

Fourth, articles were included that described studies in which the reported data could be used to examine DA's predictive validity. Studies of concurrent and construct validity were excluded (e.g., Campione, Brown, Ferrara, Jones, & Steinberg, 1985). To examine predictive validity, the analyses of included studies compared the level of performance on a DA measure to the level of performance on an achievement measure at some point in the future, or compared the level of performance on a DA measure to a future educational identification or classification. Studies that operationalized DA as an educational treatment were excluded (e.g., Feuerstein, Miller, Hoffman, Rand, Mintzker, & Jensen, 1981; Feuerstein, Rand, Hoffman, Hoffman, & Miller, 1979; Muttart, 1984; Rand et al., 1979; Savell, Twohig, & Rachford, 1986). In these studies, researchers investigated the effects of participation in a clinically-oriented DA; there were no data of a predictive nature. Finally, the operationalization of DA as different conditions of behavioral reinforcement (i.e., praise, candy, reproof) was excluded due to this criterion (e.g., Kratochwill & Severson, 1977).

Search Procedure and Identified Studies

ERIC, PsychInfo, and ECER were searched for *dynamic assessment* or *interactive assessment* or *learning potential* or *mediated assessment*. From this search, I identified the major contributors to the study of DA (e.g., Feuerstein and Budoff), and discovered a special issue of the *Journal of Special Education* devoted to the topic. In his introduction to this special issue, Haywood (1992) identified the groundbreaking research in the field of DA: Feuerstein, Rand, and Hoffman (1979); Feuerstein, Haywood, Rand, Hoffman, and Jensen (1986); Haywood and Tzuriel (1992), and Lidz (1987, 1991). In addition, two comprehensive reviews by Grigorenko and Sterberg (1998) and Swanson (2001) were read. From these resources, articles were collected that were described as studying the validity of DA or in which the title indicated that validity was studied. Finally, a second search was conducted of ERIC, PsychInfo, and ECER with the terms *dynamic assessment* or *interactive assessment* or *learning potential* or *mediated learning* and *predictive validity* to ensure that the collected studies represented most of what was available. A total of 24 studies were identified for this review. These studies are indicated by an asterisk in the Reference section.

Analysis Procedure

The data were analyzed along four dimensions. First, a comparison between traditional assessment and dynamic assessment was conducted by comparing the magnitude of the correlation coefficients measuring the association between the assessment and an achievement criterion. Second, two forms of DA were compared (one with contingent feedback and one with noncontingent feedback). Contingent feedback refers to DA that responds to students' failure with highly individualized, nonstandardized intervention. Noncontingent feedback, on the other hand, refers to DA that responds to students' failure with standardized intervention, regardless of the type of student error. Type of feedback was analyzed because it arguably speaks to the nature of classroom instruction. In classrooms with a standard approach to instruction, students would most likely receive noncontingent feedback, whereas in a classroom with a more individualized approach, students would likely receive more contingent feedback.

Third, the predictive validity of DA was analyzed across four populations: mixed ability groups, normally-achieving students, students who are at-risk or disadvantaged but not disabled, and students with disabilities. Second language learners were classified as at-risk or disadvantaged. To use DA as a tool for identification, it is especially important that the predictive validity be strong for at-risk students and students with disabilities because these groups of students are particularly susceptible to the floor effects of traditional tests discussed earlier.

Fourth, the achievement criterion was analyzed to determine whether DA could best predict (a) independent performance on the posttest of dynamic assessment measure (referred to as "posttest DA"), (b) norm-referenced achievement tests, (c) criterion-referenced achievement tests, or (d) teacher judgment. Posttest DA is the score on the DA measure given at the end of the study. It is the same measure given at the beginning of the study, but the administration is different. For posttest DA, the examiner does not offer corrective feedback to the student. The posttest DA measure represents independent student performance on identical content measured by the pretest DA.

Norm-referenced achievement tests are any commercially available assessments of achievement. Criterion-referenced achievement tests are researcher-designed assessments created with the intention of measuring the same construct as the DA administered in the study. Teacher judgment is a rating of the students' achievement in the classroom.

After analyzing the data along the four dimensions, additional analysis that explored the value added of DA, over and above traditional assessment, was investigated by finding studies in which researchers used forced entry multiple regression to investigate how much variance DA could explain after the variance due to traditional assessment was explained. If DA explains significant added variance, it may be worth the time and effort to develop new protocols and use them for identification and placement.

Mixed methods were used to explore the data. In the quantitative analysis, Pearson's correlation coefficients were used as an indicator of prediction strength. This correlation statistic served as a common metric across 15 studies. If multiple correlations were reported, the appropriate correlations were averaged to provide only one correlation statistic per analysis category per study. For example, if DA with contingent feedback was used to predict both math and reading, the two correlations were averaged to determine one correlation within the contingent vs. noncontingent analysis category. Studies in which authors did not report a Pearson's correlation coefficient were discussed descriptively. Researchers in this latter set of studies used various group and single subject designs that produced data that were not directly comparable to Pearson's correlation coefficient. Nevertheless, this information was considered valuable, because of the small number of studies exploring the predictive validity of DA.

Significance testing between average correlation coefficients was not possible due to small samples and low statistical power. Trends in the magnitude and direction of the coefficients, therefore, are discussed in lieu of statistical significance. Table 1 presents the relevant studies and corresponding correlation coefficients along the four dimensions: DA vs. traditional assessment, contingent feedback vs. noncontingent feedback, population (mixed ability groups vs. normally-achieving students vs. students who are at-risk or disadvantaged vs. students with disabilities), and achievement criterion (posttest DA vs. norm-referenced achievement tests vs. criterion-referenced achievement tests vs. teacher judgment).

Table 1

Average correlation per study within analysis categories.

Study	DA vs Traditional		Feedback		Population				Achievement Criterion			
	DA	Traditional	C	NC	Mixed	NA	AR/D	Dis	Post DA	Norm-Referenced	Criterion-Referenced	Teacher Judgment
Babad & Budoff (1974)	0.39	0.27		0.39	0.39	0.36	0.34	0.35				0.39
Bain & Olswang (1996)	–	–		–				–			–	
Bryant (1992)	0.64	0.49		0.64	0.64				0.64			
Bryant et al. (1983)	0.57	0.52		0.57	0.57				0.57			
Budoff et al.(1974)	–	–		–			–			–		
Budoff et al. (1971)	–	–	–					–			–	
Byrne et al. (2000)	–	–	–		–					–		
Day et al. (1997)	0.24	0.41	0.24			0.24			0.24			
Ferrara (1987)	0.57	0.38		0.57		0.57			0.57			
Hessels & Hamers (1993)	0.41	0.51	0.41				0.41				0.41	
Lidz et al. (1997)	0.59	0.60	0.59					0.59			0.59	
Meijer (1993)	–	–	–		–						–	
Olswang & Bain (1996)	0.73	0.33		0.73				0.73			0.73	

Table 1 Continued

Study	DA vs Traditional		Feedback		Population				Achievement Criterion			
	DA	Traditional	C	NC	Mixed	NA	AR/D	Dis	Post DA	Norm-Referenced	Criterion-Referenced	Teacher Judgment
Pena et al. (1997)	–	–	–				–			–		
Reising (1993)	–	–		–				–				–
Rutland & Campbell (1995)	0.68	0.50		0.68				0.68	0.68			
Samuels et al. (1996)	–	–	–		–					–		
Severson (1979)	0.32	0.39	0.32			0.37	0.28			0.32		
Sewell & Severson (1974)	0.41	0.41	0.41				0.41			0.41		
Spector (1992)	0.58	0.29		0.58		0.58					0.58	
Speece et al. (1990)	0.44	0.48		0.44			0.44			0.44		
Swanson (1994)	–	–	–		–					–		
Swanson (1995)	0.36	0.18	0.36		0.36					0.36		
Tissink et al. (1993)	0.46	0.35		0.46	0.46						0.46	
Average	0.49	0.41	0.39	0.56	0.46	0.42	0.37	0.59	0.53	0.38	0.63	0.39

Note: C = contingent; NC = noncontingent; Mixed = mixed ability group; NA = normally-achieving; AR/D = at-risk/disadvantaged; Dis = disability; Post DA = posttest DA score; Norm-referenced = norm-referenced achievement test; Criterion-referenced = criterion-referenced achievement test; "—" = information not reported.

Findings

DA vs. Traditional Assessment

Correlations between DA measures and achievement measures were reported in 15 of the 24 studies, and correlations between traditional assessment measures and achievement measures were also reported in the same 15 studies (Babad & Budoff, 1974; Bryant, 1982; Bryant, Brown, & Campione, 1983; Day, Englehardt, Maxwell, & Bolig, 1997; Ferrara, 1987; Hessels & Hamers, 1993; Lidz, Jepsen, & Miller, 1997; Olswang & Bain, 1996; Rutland & Campbell, 1995; Severson, 1979; Sewell & Severson, 1974; Spector, 1992; Speece, Cooper, & Kibler, 1990; Swanson, 1995; Tissink, Hamers, & Van Luit, 1993). The average correlation between DA and achievement measures was 0.49. The average correlation between traditional assessment and achievement measures was 0.41. Correlations equal to or greater than 0.40 are considered by some to be "large" (Cohen, 1977, 1988; Lipsey & Wilson, 2000). In the prediction of academic achievement, however, these correlations seem modest. Pearson's correlation coefficients do not consider the shared variance between traditional and dynamic measures, and it is impossible to determine the unique predictive ability of traditional or dynamic measures with the use of these correlations (Lipsey & Wilson, 2000).

Nine of the 24 studies investigated the predictive validity of DA without reporting Pearson's correlation coefficients (Bain & Olswang, 1995; Budoff, Gimon, & Corman, 1974; Budoff, et al., 1971; Byrne, Fielding-Barnsley, & Ashley, 2000; Meijer, 1993; Peña et al., 1992; Reising, 1993; Samuels, Killip, MacKenzie, & Fagan, 1992; Swanson, 1994). These studies were grouped into three categories according to their design and analysis: single subject design with visual analysis (Bain & Olswang, 1995), quasi-experimental design with multiple regression analysis (Budoff et al., 1974; Byrne et al., 2000; Meijer, 1993; Reising, 1993; Swanson, 1994), and experimental design with between-groups comparisons (Budoff et al., 1971; Peña et al., 1992; Samuels et al., 1992).

Single subject design with visual analysis. Bain and Olswang (1995) studied the validity of DA to predict future speech growth in a sample of 15 preschoolers with specific language impairment. Data were displayed on two scatterplots. The first scatterplot displayed participants based on their weighted DA score for both semantic and functional relations against their change in mean length utterance (MLU) during the nine week study. Results indicated that the weighted DA score accurately predicted change in rate of learning for 12 of the 15 participants. The second graph plotted participants' weighted DA score for only semantic relations against their change in MLU. Results indicated that the weighted DA score accurately predicted the change in rate of learning for all 15 participants. That is, those with the highest weighted DA score showed the greatest gains in speech.

Quasi-experimental design with multiple regression analysis. Budoff et al. (1974), Byrne et al. (2000), Meijer (1993), Reising (1993), and Swanson (1994) used multiple regression analyses to study the unique predictive ability of DA over and above traditional assessment. All studies used some form of verbal and quantitative achievement as criteria to determine predictive validity. Budoff et al. found mixed results with a population of disadvantaged students: DA was significantly better than traditional assessment in the prediction of nonverbal/quantitative achievement; however, patterns of prediction for verbal measures were inconsistent. Although DA scores were a statistically significant predictor of one of the four verbal measures, traditional measures (e.g., IQ) and demographic information (e.g., age) were generally more consistent predictors.

By contrast, Byrne et al. (2000), Meijer (1993), and Reising (1993) showed that DA made a significant and consistent contribution to the prediction of achievement. Byrne et al. used a DA procedure called session of last error to predict future phonemic awareness and reading achievement. Session of last error is a measure of the rate of reading progress throughout the study. It is closer to the current operationalization of RTI than the more compact notion of DA because it tracks student achievement for several weeks. The faster students reached mastery, the earlier their session of last error.

Byrne et al. (2000) studied the reading achievement of a cohort of children in kindergarten and conducted follow-up tests in second and fifth grade. Byrne and his colleagues performed a series of multiple regression analyses on achievement in kindergarten, second grade, and fifth grade. In each of the analyses, the posttest traditional score was entered first into the equation. Session of least error was entered as the second predictive variable. In all cases, the session of least error, was a significant predictor of achievement above and beyond the traditional posttest score. It explained from 9% to 21% of the total variance.

Meijer (1993) performed a similar analysis on math achievement of a mixed-ability group of secondary students. First, a traditional measure of initial math achievement was entered into the multiple regression, which accounted for 11% of the variance in achievement. Second, a DA measure was added as a predictor, and it accounted for an additional 13% of the variance. Similarly, Reising (1993) found that, after controlling for verbal IQ, the combination of two dynamic measures (number of hints required to solve a problem and number of items requiring help) predicted an additional 13% of the variance in verbal achievement, 18% of the variance in math achievement, and 14% of the variance in teacher ratings of school performance for primary students with disabilities.

Swanson (1994) conducted two separate multiple regression analyses on a mixed-ability group of primary students. In the first analysis, the initial traditional score was entered before dynamic variables. For reading achievement, the initial traditional score explained 11% of the total variance and a combination of dynamic scores explained an additional 19%. For math achievement, the initial traditional score explained 20% of the total variance and a processing stability score (initial score minus maintenance score) explained an additional 12%. DA did not explain unique variance in math achievement. In the second regression analysis, all variables were allowed to compete against each other. For reading achievement, three DA measures (gain score, probe score, and maintenance score) were found to be the best predictors of achievement, explaining a total of 34% of the variance. For math achievement, only one DA measure (gain score) was a significant predictor of achievement, explaining 32% of the variance. The ability of DA to predict future achievement, therefore, may depend on what domain of achievement is being predicted and whether initial traditional scores are entered as the first variable in a multiple regression.

Experimental design with between group comparisons. Three studies investigated the predictive validity of DA with experimental methods (Budoff et al., 1971; Peña et al., 1992; Samuels et al., 1992). Budoff et al. studied DA's utility in predicting the response to a classroom science curriculum for low-achieving students in grades 7 through 9. Even after IQ was factored out, performance on DA predicted which students would respond positively to the science curriculum, $F(9,39) = 4.17$, $p < .001$. That is, students who initially scored higher on DA *or* students who improved throughout the administration of DA tended to learn more than students who scored lower on DA and showed no improvement during its administration.

Peña et al. (1992) used DA to differentiate Spanish-speaking preschool students with language disorders from nondisabled Spanish-speaking students who had poor English skills. Peña and her colleagues developed a measure of learning potential called the modifiability index. Results indicated that language-disordered students had a significantly lower modifiability index than nondisabled students, $F(1,36) = 53.21$, $p < .00001$. Additionally, students with a higher modifiability index demonstrated more gain in single word vocabulary over the course of the school year, $F(1,46) = 13.52$, $p = .0006$. Peña et al. concluded that static measures alone would over identify Spanish-speaking students for special education placements, but DA demonstrated the ability to assess learning potential and to differentiate students with language disorders from nondisabled students.

Another potential use of DA is informing educational placement. Samuels et al. (1992) studied DA in regards to its prediction of regular versus special education placement of students after preschool. DA significantly predicted educational placement (regular versus special), $\chi^2(2) = 6.48$, $p<.05$. Results also indicated that placement could not be predicted on the basis of a traditional receptive vocabulary measure (Peabody Picture Vocabulary Test, Revised). Samuels et al. concluded that traditional assessment alone could not fully capture the potential of a student, and that DA may be an important tool for placement and programming decisions.

Summary. DA and traditional assessments correlate similarly to future achievement measures. Beyond traditional assessments, however, researchers of DA have demonstrated that DA can identify students who will respond to instruction (Bain & Olswang, 1995, Budoff et al., 1971), differentiate minority students with and without language disorders (Peña et al., 1992), and predict future educational placement (Samuels et al., 1992). Some studies have reported that DA can contribute to the prediction of achievement beyond traditional assessments (Byrne et al., 2000; Meijer, 1993; Reising, 1993). Results are inconsistent and sometimes depend on analysis techniques and domains of study (Swanson, 1994).

Feedback: Does the Type of Feedback in Dynamic Assessment Affect Predictive Validity?

Of the 15 DA studies reporting Pearson's correlation coefficients, 6 provided contingent feedback (individualized instruction in response to student failure) and 9 provided noncontingent feedback (standardized instruction in response to student failure). Studies with contingent feedback correlated 0.39 with achievement, whereas studies with noncontingent feedback correlated 0.56 with achievement. Nine studies did not report Pearson's correlation coefficients: 6 studies with contingent feedback (Budoff et al., 1974; Byrne et al., 2000; Meijer, 1993; Peña et al., 1992; Samuels et al., 1992; Swanson, 1994) and 3 studies with noncontingent feedback (Bain & Olswang, 1995; Budoff et al., 1971; Reising, 1993).

Contingent feedback. It was difficult to investigate contingent feedback studies as a group (n=6) because the study designs operationalized achievement variables in different ways (continuous or dichotomous), which changed the meaning of “significant” results. When achievement was operationalized as a continuous variable (i.e., an achievement test), two studies reported positive support for the predictive validity of DA (Budoff et al., 1974; Byrne et al., 2000), and two additional studies reported mixed findings (Meijer, 1993; Swanson, 1994) such that results depended on the analysis technique and achievement domain in question. Two other studies operationalized achievement as a dichotomous variable and found that DA can predict identification or educational placement (Peña et al., 1992; Samuels et al., 1992). When an inherently continuous variable (i.e., achievement) is transformed into an artificial dichotomy (i.e., educational placement using an achievement cut-off point), statistical significance is not equivalent to the statistical significance obtained with a continuous variable (Lipsey & Wilson, 2000). That is, significance is numerically easier to obtain with dichotomous variables than with continuous variables.

Noncontingent feedback. The results of the studies using noncontingent feedback were somewhat more straightforward. Using visual analysis, Bain and Olswang (1995) found that their noncontingent DA measure predicted immediate growth in speech with consistency. In addition, Budoff et al. (1971) and Reising (1993) found that DA predicted unique variance above and beyond that which was predicted by IQ.

Summary. Trends in Pearson’s correlation coefficients show that DA with noncontingent feedback is more strongly associated with future achievement than DA with contingent feedback. Studies using contingent feedback that do not report correlation coefficients are difficult to synthesize across participants and across studies because of their highly individualized nature. Studies using noncontingent feedback that do not report correlation coefficients are somewhat easier to synthesize and generally provide evidence that DA is useful in the prediction of future achievement, even when used in conjunction with traditional assessments.

Population: For Whom Does Dynamic Assessment Have Predictive Validity?

Study participants were separated into four categories: mixed ability groups, normally-achieving students, at-risk or disadvantaged students, and students with disabilities. Some studies reported data separately for more than one participant group, and therefore provided Pearson's correlation coefficients in more than one category. Correlations were provided for 5 studies with mixed ability groups ($r = 0.46$), 5 studies with normally-achieving students ($r = 0.42$), 5 studies with at-risk or disadvantaged students ($r = 0.37$), and 4 studies with students with disabilities ($r = 0.59$).

Normally-achieving students. All of the studies with normally-achieving students provided Pearson's correlation coefficients. DA correlated 0.42 with outcome measures.

Mixed-ability groups. Four studies with mixed ability groups did not provide Pearson's correlation coefficients. These results will not be discussed because they do not differentiate normally-achieving students from at-risk students from students with disabilities. The data in mixed ability group studies were not disaggregated by population. With no details on the mixed ability group, it is impossible to tell what type of student (i.e., normally-achieving, at-risk, or disabled) contributed most significantly to the results.

At-risk students. Achievement of at-risk or disadvantaged students, for whom DA measures are often designed, is predicted with slightly less accuracy than for mixed ability groups and normally-achieving students. Two studies with at-risk or disadvantaged students did not report Pearson's correlation coefficients (Budoff et al., 1974; Peña et al., 1992). As discussed, Budoff et al. (1974) found that DA scores were significant, yet inconsistent, predictors of achievement. The results of Peña et al. indicated that DA can differentiate disabled from nondisabled Spanish-speaking preschool children and predict English language growth.

Students with disabilities. DA predicted the academic achievement of students with disabilities with slightly more accuracy than the other three groups. Two DA studies predicting the achievement of students with disabilities did not provide Pearson's correlation coefficients: Bain & Olswang (1995) and Budoff et al. (1971). The results of these two studies, as discussed, support the quantitative trend of correlation coefficients indicating that DA may be a better predictor of achievement than traditional assessment for students with disabilities.

Summary. Trends in correlation coefficients show that DA was most strongly correlated with achievement for students with disabilities. The correlation between DA and achievement was weakest for at-risk or disadvantaged students. Ironically, DA is often designed to create a less biased measure of achievement for at-risk students. These results indicate that DA may not be less biased than traditional assessment for this population.

Achievement Criterion: What Achievement Measures Can Dynamic Assessment Predict?

There were four types of achievement criteria: independent performance on the posttest DA measure (posttest DA), norm-referenced achievement tests, criterion-referenced achievement tests, and teacher judgment. Posttest DA is the achievement measure that is most similar to the DA measure itself. In most cases, the posttest DA is simply an alternate form of the pretest and training phases of DA. Criterion-referenced achievement tests are the next most similar to the DA measure. These criterion-referenced achievement tests are designed by the researcher to measure the same construct being taught during the DA. Norm-referenced achievement tests, by contrast, may or may not be similar to the DA measure.

Fifteen studies provided Pearson's correlation coefficients: 5 predicted posttest DA, 4 predicted norm-referenced achievement tests, 5 predicted criterion-referenced achievement tests, and 1 predicted teacher judgment. DA measures correlated 0.53 with posttest DA, 0.38 with norm-referenced achievement tests, 0.63 with criterion-referenced achievement tests, and 0.39 with teacher judgment. The trend of the correlations is interesting with respect to the similarity of the DA measure to the achievement measure. Measures more similar to DA, such as posttest DA and criterion-referenced achievement tests, are predicted with greater accuracy (0.53 and 0.63 respectively) than those measures that are less similar, such as norm-referenced achievement tests and teacher judgment (0.38 and 0.39 respectively).

Posttest DA. All studies that predicted posttest DA provided Pearson's correlation coefficients. DA measures correlated 0.53 with independent posttest DA performance.

Norm-referenced achievement tests. Five studies that predicted norm-referenced achievement tests did not provide correlation coefficients (Budoff et al., 1974; Byrne et al., 2000; Peña et al., 1992; Samuels et al., 1992; Swanson, 1994). Mixed support was found for DA's ability to predict achievement as measured by norm-referenced tests. As discussed, Peña et al. (1992) and Samuels et al. (1992) found positive support for the use of DA as a tool for identification and placement, respectively, and Byrne et al. (2000) determined that DA explained unique variance in achievement. Budoff et al., (1974) and Swanson (1994) found mixed results. Demographic factors and traditional assessment were more consistent predictors than DA in Budoff et al.'s study; and Swanson found that the significance of the results depended on analysis techniques and the academic domain in question.

Criterion-referenced achievement tests. Four studies that predicted criterion-referenced achievement did not provide correlation coefficients (Bain & Olswang, 1995; Budoff et al., 1971; Meijer, 1993; Reising, 1993). As discussed, Bain and Olswang (1995) and Budoff et al., (1971) found positive support for the ability of DA to predict growth in achievement. Meijer (1993) and Reising (1993) both concluded that DA explained unique variance in the prediction of achievement, even after intelligence had been factored out. DA was a consistently significant predictor in the prediction of achievement as measured by criterion-referenced tests.

Teacher judgment. One study that predicted teacher judgment (Reising, 1993) did not report Pearson's correlation coefficients. Although DA did not predict teacher judgment as well as posttest DA or criterion-referenced achievement tests, one study (Reising, 1993) found that DA accounted for 14% of the variance in teacher judgment of achievement, even after IQ had been factored out.

Summary. Again, the studies that did not report Pearson's correlation coefficients seemed to generally follow the trend of the quantitative analysis. Posttest DA and criterion-referenced achievement tests were predicted more consistently than norm-referenced achievement tests and teacher judgment

Discussion

The purpose of this review was to synthesize evidence on the predictive validity of DA. Pearson correlation coefficients indicated that traditional and dynamic assessments predict future achievement with similar accuracy. Trends among the correlation coefficients indicated that DA predicted achievement more accurately (a) when the feedback of the assessment was noncontingent on the student response, (b) with respect to the achievement of students with disabilities, rather than at-risk or normally-achieving students, and (c) when involving independent DA posttests and criterion-referenced achievement tests instead of norm-referenced achievement tests and teacher judgment of student achievement.

If traditional and dynamic assessments do equally well in predicting achievement, why should we consider using DA? If DA is time consuming to develop and validate, why exert the extra effort to develop new tests when valid traditional assessments are already available?

To address this question, we must consider another question: Whether traditional assessment and DA are measuring the same constructs that predict achievement. Past reviews have not focused on whether DA explains unique variance in student achievement. To examine this, we must look at the value added of DA over and above traditional assessment. This is possible in analyses in which researchers used forced entry multiple regression. If traditional variables are entered first, it is possible to examine DA's unique contribution to the variance in achievement.

Does DA Provide Added Value to Traditional Assessment?

Ten studies conducted a forced entry multiple regression analysis to explore DA's unique ability to predict achievement over and above traditional assessment (Bryant 1982; Bryant et al., 1983; Byrne et al., 2000; Ferrara, 1987; Meijer, 1993; Reising, 1993; Rutland & Campbell, 1995; Spector, 1992; Speece et al., 1990; Tissink et al., 1993). Two studies (Byrne et al., 2000 and Meijer, 1993) investigated the unique contribution of DA after traditional *achievement* tests had been entered in the multiple regression, and eight studies investigated the unique contribution of DA after traditional *cognitive* tests (i.e., IQ tests) had been entered in the multiple regression.

Value added to traditional achievement tests. DA consistently contributed significant unique variance to the prediction of future achievement above and beyond traditional achievement tests. Byrne et al. (2000) found that DA accounted for an additional 9% to 21% of the variance in phonemic awareness and reading achievement for students in kindergarten, grade 2, and grade 5. Likewise, Meijer (1993) found that DA accounted for an additional 13% of the variance in math achievement for secondary students.

Value added to traditional cognitive tests. DA also consistently contributed significant unique variance to the prediction of future achievement above and beyond traditional cognitive tests. The eight studies in which researchers conducted these analyses predicted three domains: general reasoning, verbal achievement, and math achievement. In the domain of general reasoning, researchers investigated student performance on measures such as mazes, matrices, and series completion. Bryant (1982) found that two DA measures predicted significant variance in achievement: training score (22%) and transfer score (17%). Similarly, Bryant, Brown, and Campione (1983) found that transfer score explained 22% of the variance in achievement above and beyond IQ (although the training score was found to be nonsignificant). Rutland and Campbell (1995) found that dynamic training, maintenance, and transfer all made significant contributions to the variance in achievement (11%, 11%, and 9%, respectively).

In the verbal domain, DA also consistently contributed to the prediction of achievement. Spector (1992) found that DA contributed between 12% and 14% on phonological awareness measures and 21% on a word reading measure. Indeed, DA was the only significant predictor of word reading. Reising (1993) stated that DA contributed an additional 13% in higher-level verbal measures, such as reading sentences and writing. Speece et al. (1990), however, reported that DA was not a significant predictor of verbal achievement. The only significant predictors of verbal achievement in this study were verbal IQ and traditional pre-test (25% combined).

Results concerning the added value of DA in the prediction of math achievement were consistent, although they varied greatly in magnitude. Ferrara (1987) noted that two dynamic measures explained a statistically significant portion of the variance in math growth: training score (17%) and maintenance and transfer score (32%). Reising (1993) and Tissink et al. (1993) also found that DA contributed significant variance to math achievement although it contributed less so than Ferrara’s study (18% and 7% respectively). Speece et al. (1990) reported that DA training contributed significant variance to math achievement; however, it explained only 2% of the overall variance.

In general, there is evidence that DA can predict unique achievement that is not tapped by traditional achievement or traditional cognitive assessment. When DA scores were entered after traditional scores in a forced entry multiple regression, they explained significant variance in the prediction of general reasoning, verbal achievement, and math achievement. Only one study (Speece et al., 1990) was inconsistent with these results. Future research, therefore, must acknowledge that DA may not be a substitute for traditional assessment. Rather, it may provide valuable information over and above that which traditional assessment provides. The practical significance of this additional information, however, is not yet understood.

Limitations of Review

There are very few quantitative syntheses of DA research (e.g., Swanson, 2001) and none that are concerned primarily with predictive validity. It is difficult to synthesize research on such a broad and sometimes poorly defined topic, and these results must be understood relative to the paucity of studies.

Nature of the study design. Several studies in this review were not primarily concerned with measuring the predictive validity of DA. DA measures may not have been designed with the specific purpose of prediction and identification. Similarly, the achievement measures may not have been chosen specifically to measure change across time. In addition, both the DA measures and criterion-referenced achievement measures had unreported psychometric properties. We cannot be sure that the constructs that were measured were valid, that the measures were reliable, or that the measures were implemented with fidelity.

Study rigor. One final note concerns the relationship of DA feedback and study rigor. In well-controlled research, the researcher strives to minimize variables that will confound results. It is easier to conduct rigorous research in DA using standardized, noncontingent feedback. Individualized, contingent feedback is more difficult to control. Researchers using noncontingent feedback may be exploring performance using methods that are easier to measure, quantify, and analyze. In such studies, standardized procedures are used in all cases of student failure; therefore, the independent variable is clear and unchanging. Researchers using contingent feedback, by contrast, introduce an "if/then" process into intervention. For example, if the students fail because they did not understand the directions, then the teacher may need to repeat or clarify the directions. If the students fail because they lack the underlying skills necessary for success, then the teacher may need to concentrate on teaching lower-level skills. How can we compare the results of DA across students who require individualized intervention? If the instructional elements are not the same, how can we determine that the predictive ability is due to the nature of the DA and not to the teacher, teaching method, or some other unmeasured variable? It may be that noncontingent and contingent feedback cannot be judged by the same standards of rigor. And, consequently, it may not be appropriate to compare noncontingent and contingent feedback using current research methods because noncontingent feedback fits more easily into the framework of rigorous, empirical research and therefore, is more likely to produce consistent results. Clinically-oriented DA that uses contingent feedback may need to develop new and different standards of rigor.

Relevance to Present Study

This review has summarized research on the predictive validity of DA in comparison to traditional assessments. To make the issue of DA more current, this study investigates the predictive validity of DA in comparison to progress monitoring within an RTI framework. DA vs. RTI is an interesting comparison because of their conceptual similarities. Both approaches measure independent and assisted performance, and both approaches consider “unresponsiveness” a necessary (though not sufficient) condition for special education services. For the purposes of identification and placement, the main difference between DA and RTI approaches is the timeline of assessment. DA is designed to measure learning ease within one testing session, whereas RTI approaches use multiple testing sessions across several weeks or months. If DA can be equally effective as RTI in identifying students who are at-risk for school failure, we have the potential to provide more appropriate intervention to “nonresponders” at an earlier date.

CHAPTER III

METHOD

Participants

Schools. This study took place as part of a larger study investigating the psychometric properties of a DA measure designed by Fuchs, Fuchs, and Compton (2004). Four schools from the Metropolitan-Nashville Public Schools were recruited to participate. Two of the four schools received Title 1 funding.

Teachers. Ten kindergarten teachers and twelve first grade teachers agreed to participate. The 22 teachers permitted examiners to pull students from their classrooms, and they completed questionnaires and surveys on student demographics and attention. In return for their cooperation, the teachers were given cash stipends. Table 2 presents demographic information on the teachers.

Students. A total of 233 students consented to participate. Seventeen students did not participate in the screening due to delayed parental consent or excessive absences. A total of 216 children were screened and 133 students (28 in kindergarten and 105 in first grade) were selected to participate in the remainder of the study. Only higher performing kindergarten students were selected due to the difficulty of the DA measure. All 105 screened first grade students were selected to participate.

Eight first grade students were removed from the sample due to invalid pretest data. Five additional students changed schools and were no longer reachable during the study. The final sample consisted of 120 students; 25 at kindergarten and 95 at first grade. Table 3 presents student demographic information on the final sample. The kindergarten and first grade samples are not comparable. The first grade sample differed from the kindergarten sample in that there were higher percentages of minority students, students who received free or reduced lunch, students with IEPs, and students who had previously been retained.

Table 2

Teacher Demographics

Total Teachers	22
Females	22
Race	
African-American	4
Caucasian	17
Other	1
Age	
20-29	4
30-39	5
40-49	6
50-59	4
60-69	3
Median highest Degree earned	MEd/MS
Mean years teaching experience	14.82 (9.75)
Mean years in current position	9.5 (7.34)
Number of teachers in Title 1 schools	10
Number of credit hours in reading	
0-3	4
4-6	5
7-12	4
13+	9
Number of credit hours in special education	
0-3	13
4-6	4
7-12	2
13+	3

Note: SD in parentheses.

Table 3

Student Demographics

Total Students	120
Kindergarten	25
Gender	
Female	12 (48%)
Male	13 (52%)
Race	
African-American	5 (20%)
Caucasian	15 (60%)
Hispanic	2 (8%)
Asian	1 (4%)
Other	2 (8%)
Number of students receiving free or reduced lunch	11 (44%)
Number of students with IEP	0 (0%)
Number of students previously retained	1 (4%)
First Grade	95
Gender	
Female	41 (43%)
Male	54 (57%)
Race	
African-American	50 (53%)
Caucasian	30 (32%)
Hispanic	6 (6%)
Asian	3 (3%)
Other	6 (6%)
Number of students receiving free or reduced lunch	64 (67%)
Number of students with IEP	8 (8%)
Number of students previously retained	7 (7%)

Note: Percentages in parentheses (separate percentages for kindergarten and first grade). Percentages within categories may not total 100% due to rounding error.

Procedures

Examiner training. One project coordinator and nine research assistants conducted the student assessments. The research assistants were either masters or doctoral degree students. They received extensive test training, which included (a) modeling by the project coordinator and a doctoral research assistant (20 hours), (b) listening to tapes from testing sessions to practice scoring (10 hours), (c) role playing with other research assistants (35 hours), and (d) independent practice (5 hours).

Examiner fidelity. Fidelity of test implementation was obtained for each research assistant. If they did not reach criterion (i.e., correctly performing at least 90% of the testing procedures), they were given additional training and tested again. The fidelity checklist for traditional assessment can be found in Appendix A, and the fidelity checklist for DA can be found in Appendix B. In addition, inter-rater agreement was measured between the project coordinator and all research assistants. If the research assistants did not reach criterion (i.e., 90% of scored items were identical to the project coordinator's scored items), they were given additional training and tested again.

Measures: Screening for Study Selection and Traditional Battery

The *Letter Sounds* screening measure consists of 30 sounds: 21 consonants, 5 vowels, and 5 blends/letter combinations (qu, sh, ch, th, ck). Students are directed to do their best to say the sound the letter makes. The measure is untimed, but if students hesitate for more than 5 seconds, they are directed to move on. The scores range from 0 to 30. The *Decoding Inventory* consists of 20 decodable nonwords: 5 CVC, 5 CVCE, and 10 words that had a single or double consonant and the suffix –ing (referred to as Doubling; e.g., loting, mutting). The students are guided through two sample items (mim, op). Then they are shown the 20 nonwords and instructed to tell the examiner how the words sound. The scores range from 0 to 5 for CVC, 0 to 5 for CVCE, and 0 to 10 for Doubling.

The *WRMT-R Word Attack* test is a measure of phonetic reading ability (Torgeson, Wagner, & Rashotte, 1997). The test consists of 45 items arranged in order of difficulty. The test is discontinued when a student answers six consecutive items incorrectly, or when all 45 items have been administered. The score ranges from 0 to 45. The internal consistency for 1st grade students ranges from 0.94 to 0.97.

The *WRAT Reading* subtest consists of two parts. In the first part, students are required to name 15 letters. In the second part, students are instructed to do as well as they can to read single words. The test is discontinued when the student answers 10 consecutive items incorrectly (letter, words, or letters and words together). Scores range from 0 to 57. The manual reports split-half reliability of 0.98 for *WRAT Reading*.

The *Fluency* subtest consisted of two decodable short stories: *Jim and the Pet Pig* and *The Cat and the Dog*. Students are given 60-seconds for each story and instructed to read as quickly and correctly as possible. Raw scores range from 0 to 64 on the first fluency measure and 0 to 74 on the second. Scores are adjusted if the student finishes in fewer than 60 seconds. The final Fluency score was the average words read per minute of the two stories.

The *WIAT Spelling* subtest measures students' abilities to write letters and words. Items 1 through 4 ask the student to reproduce letters; items 5 and 6 as the student to reproduce sounds; and items 7-50 ask the student to reproduce words. Students are given 10 seconds for each item. Raw scores range from 0 to 50.

Measures: Dynamic Assessment Battery

The Dynamic Assessment (DA) measure was designed by Fuchs, Fuchs, and Compton (2004). It consists of nonwords separated into three subtests: CVC, CVCE, and Doubling Consonant. In general, each subtest requires the child to learn a decoding "rule" (i.e., short vs. long vowels). All nonwords have either a short "o" or long "o" vowel sound. In each subtest, students are given five opportunities (represented by levels) to master the content. At any particular level, if students read 5 of 6 words correctly, they are regarded as having mastered the skill. If students fail to master the content at Level 1, they are given a hint to help them learn the decoding rule (i.e., CVC, CVCE, or Doubling Consonant). If students fail to master the content at Level 2, they are given a more explicit hint. Increasingly explicit hints are given until the student reaches mastery or until all hints have been administered (Level 5).

If students do not reach mastery by Level 5 of the CVC subtest, the CVCE and Doubling Consonant subtests are not administered. Similarly, if students do not reach mastery on the CVCE subtest, the Doubling subtest is not administered. Each subtest is scored 1 through 5. A score of 1 indicates that a student reached mastery at the first opportunity (Level 1); a score of 5 indicates that a student reached mastery at the fifth and final opportunity (Level 5). In other words, a lower score indicates quicker mastery of content. If students are not administered a subtest due to lack of mastery of lower content, they are automatically given a score of 5. Thus, the best total DA score is 3; the poorest score is 15.

CVC. For the CVC subtest, the nonword test items at each level are *fot, gop, vop, wot, jop,* and *zot*. *Level 1: Reading to the Child* includes modeling the reading of nonsense words with the short “o” sound (e.g., *bod* and *zod*). *Level 2: Teaching Onset* instructs the student to attend to the first sound of each nonsense word. *Level 3: Teaching Rime* instructs the student to attend to the last two sounds of each word. *Level 4: Teaching Onset-Rime Blending I* instructs the student to decode the onset and rime separately and then to blend them together into a word. *Level 5: Teaching Onset-Rime Blending II* teaches the same content as *Level 4* along with the examiner explicitly stating the decoding rule.

CVCE. The nonword test items at each level of the CVCE subtest are *fote, gope, vope, wote, jope,* and *zote*. *Level 1: Reading to the Child* includes modeling reading of nonsense words with the short “o” and long “o” sounds (e.g., *bod* and *bode*). *Level 2: Hearing Long and Short Middle Vowel Sounds* instructs the student to listen to the difference between the short “o” sound and long “o” sound in word pairs (e.g., *dod* and *dode*). *Level 3: Teaching “Long” and “Short” Vowel Terminology* instructs students to use the terms “long ‘o’” and “short ‘o’” and to recognize their visual symbols (i.e., “ō” and “ŏ*”). Level 4: Teaching the “Magic e” Rule* instructs the student that when there is an “e” at the end of the word, the “o” says its name and makes the long “o” sound; and, when there is no “e” at the end of the word, the “o” does not say its name and makes the short “o” sound. *Level 5: Teaching the “Magic e” Rule with Color Emphasis* is identical to *Level 4,* except the “Magic e” is colored red to help the student attend to it.

Doubling Consonant. In the Doubling Consonant subtest, the nonword test items at each level are *fotting, goping, vopping, woting, jopping,* and *zoting*. Before any of the testing levels are administered, the examiner conducts a “pre-teaching” session to make sure the student can recognize “—ing” and say its sound /ing/. *Level 1: Reading to the*

Child includes modeling nonsense words with single and double consonants that also have the suffix –ing (e.g., *boding* and *bodding*). *Level 2: Long vs. Short Vowel Sound* instructs students to listen to the number of sounds in each word and determine if the vowel sounds the same or different. In *Level 3: Single vs. Double Consonant* the examiner models words while students are told to attend to whether the word has a single or double consonant; however, no explicit rule is stated. *Level 4: Teaching the Doubling Rule* instructs the student that when a word has one consonant, the "o" says its name and makes the long "o" sound. When a word has two consonants, the "o" does not say its name and makes the short "o" sound. *Level 5: Teaching the Doubling Rule with Color Emphasis* is identical to Level 4 except that the consonant in single consonant words is colored red and the consonants in the double consonant words are colored green to help the student attend to the difference.

Measures: Curriculum-Based Measurement (CBM)

Two forms of CBM were used to monitor student progress: letter sound lists and word lists. The letter sound list consists of 30 sounds: 21 consonants, 5 vowels, and 4 consonant blends/clusters (qu, sh, ch, th, ck). Predictive validity of the letter sound list was studied relative to WRMT Word Identification, WRMT Word Attack, and WIAT Spelling measures (0.71, 0.66, 0.71, respectively). Test-retest reliability is reported as 0.89 and alternate-form reliability is reported as 0.94. Two forms of the CBM word lists were used. For week #1 through week #9, the word lists consisted of 50 high-frequency words. For week #10 through week #12, the word lists consisted of 100 high-frequency words. Test-retest reliability for two consecutive weeks is reported as 0.97 and for two consecutive months is 0.91 (Fuchs, Fuchs, Compton, & Bryant, 2004).

Kindergarten students were administered letter sound lists once per week from the Fall assessment to week #6. At week #7, kindergarten students were administered both letter sound lists and word lists. First grade students were administered word lists once per week for 11 weeks between Fall and Spring assessments.

Data Collection

All testing sessions were recorded with audiotape. Data were collected in four phases: screening for study selection, Fall assessment, CBM, and Spring assessment. A specific treatment was not conducted between Fall and Spring assessments. In this study, “treatment” refers to typical classroom instruction in reading that was conducted by the students’ teachers during the 11-week interval between Fall and Spring assessments.

First, an initial screening to select the study sample was conducted in November. Next, students were assessed with the Fall traditional static battery in December and the Fall DA in January. The Fall traditional battery for the larger study included RLN, Segmentation, WRMT-R Word Attack, WRAT Arithmetic, WRAT Reading, WASI Block Design, WASI Matrix Reasoning, and CBM. The order of test administration was randomized for each student.

Third, CBM was collected weekly from mid-January to mid-April (12 weeks). Finally, students were administered the Spring traditional battery in April and the Spring DA in May. The Spring traditional battery for the larger study differed from the Fall battery in three ways. First, WRAT Arithmetic was eliminated. Second, WASI Vocabulary and Similarities were administered instead of the WASI Block Design and Matrix Reasoning subtests. Third, two measures were added: oral reading fluency and the WIAT Spelling subtest. Again, order of administration of these measures was randomized for each student. Spring DA was identical to Fall DA.

Data Scoring and Data Entry

Tests were initially scored by the examiner who administered them. Scoring was double checked by either me or the project coordinator. If there were any questions in scoring, audiotapes were checked. Pairs of research assistants entered data into two independent databases: an original and a duplicate. The original and duplicate databases were compared for accuracy and modified until discrepancies were eliminated.

Inter-rater agreement. Inter-rater agreement was calculated for 15% of testing sessions from the larger study (32 students for screening and 19 students for Fall traditional, Fall DA, Spring traditional, and Spring DA). I listened to audiotapes of all subtests of screening and DA. For each Fall and Spring traditional assessment, the project coordinator randomly selected three tests for me to rescore (WASI Block Design, WASI Matrix Reasoning, and WIAT Spelling were excluded from the random selection because student responses could not be recorded using audiotapes). I rescored the subtests independently without knowledge of the original scoring. The project coordinator calculated point-by-point agreement between the original and rescored testing protocols. Inter-rater agreement is presented in Table 4 (only measures used in the present analysis are included in the following table).

Table 4

Inter-rater Agreement

Screening for Study Selection	
Letter Sounds	97%
Decoding Inventory	92%
Fall Traditional Assessment	
Word Attack	96%
WRAT Reading	98%
CBM	95%
Fall DA	100%
Spring Traditional Assessment	
Word Attack	89%
WRAT Reading	99%
Fluency	99%
CBM	96%
Spring DA	100%

Inter-rater agreement was not calculated for weekly CBM. Although students were directed to read words in order, some students (especially students with lower reading ability) skipped words without a verbal marker recorded on audiotape. After skipping any number of words, lower students often pointed to a word (e.g., “for”) and read the word incorrectly (e.g., “from”). Because the word “from” was also an item on the word list, it is difficult to determine from audiotape which word the student was attempting. Only the actual “online” examiners could score those items correctly because only they could see the student pointing.

Inter-rater agreement was calculated for CBM on the Fall and Spring Traditional Assessments. If I could not follow the student on the audiotape because of excessive word skipping, another student’s test was chosen at random and scored.

CHAPTER IV

RESULTS

First, descriptive statistics for achievement measures are reported. Then, results of a multiple regression analysis exploring possible predictors of achievement are described. There are three "screening" predictor variables used in the analysis. The first variable is Fall DA. The second and third variables are CBM intercept and CBM slope, both derived from the progress monitoring data. Outcome measures include WRAT Reading, WRMT-R Word Attack, Fluency, and WIAT Spelling. Last, a commonality analysis for the predictor variables is reported for each of the four dependent variables.

For this analysis, CBM intercept is defined as initial single word reading score at week #1. Performance at week #1 was selected as an initial performance measure, similar to that which a classroom teacher might use to predict future achievement in the classroom. CBM slope is defined as the slope of the best-fit line across 5 weeks of CBM data. Again, CBM slope was conceptualized in this way because it mirrors how slope might be calculated by classroom teachers.

Descriptive Statistics

Means and standard deviations are reported for the achievement measures in Table 5. Data are reported on the 120 participants who completed the study. Several trends require comment. First, the Decoding Inventory screening measure was subject to floor effects. The means are close to zero and, with the exception of CVC, the standard deviations are larger than the means for both kindergarten and first grade students. Second, when inspecting the data, it is important to remember that the CBM scores of the kindergarten students were different from CBM of first grade students in the Fall. In the Fall, kindergarten students on average correctly named 34.34 sounds per minute, whereas first grade students on average correctly named 18.32 words per minute. Although the kindergarten score is higher, they did not outperform first grade students because the kindergarten students were tested on lower level skills. In the Spring, both kindergarten and first grade students were tested on the number or words read correctly per minute. CBM word scores in the Spring can be legitimately compared between age groups.

Finally, the mean DA score is higher for kindergarten students than first grade students. Lower DA scores indicate that participants required *less* assistance to master reading skills. If participants improved from Fall to Spring, their DA score would decrease.

Multiple Regression Analysis

The multiple regression analysis was run with three predictor variables (Fall DA score, CBM intercept, and CBM slope) and four outcome variables (WRAT Reading, Word Attack, Fluency, and WIAT Spelling). First, Pearson correlations were calculated between the seven measures. Then, separate regression analyses were conducted for kindergarten and first grade participants. Separate analyses were required because, as indicated, the CBM intercept and CBM slope terms for kindergarten and first grade students were not comparable. The CBM slope term used as a predictor variable in the kindergarten analysis represents average weekly growth in the number of sounds named correctly per minute, whereas the CBM slope term in the first grade analysis represents average weekly growth in the number of words named correctly per minute. The CBM intercept represents students' CBM score at week #1. For kindergarten students, the CBM intercept score is the number of sounds named correctly in one minute at week #1, and for first grade students, the CBM intercept score is the number of words named correctly in one minute at week #1. (CBM intercept at week #1 will be referred to as "CBM intercept" in all future references).

Table 5.

Means and Standard Deviations of Screening, Pretest, and Posttest.

Measure	Grade: Kindergarten N=25	Grade: First Grade N=95
Screening		
Letter Sounds	25.88 (2.26)	27.15 (2.60)
Decoding Inventory – CVC	2.60 (1.47)	2.75 (1.77)
Decoding Inventory – CVCE	0.84 (1.68)	1.08 (1.57)
Decoding Inventory – Doubling	0.52 (1.53)	1.71 (2.12)
Fall Traditional Assessment		
Word Attack	6.80 (6.47)	10.89 (7.79)
WRAT Reading	19.32 (4.44)	22.05 (4.99)
CBM Sounds	34.34 (11.20)	
CBM Words		18.32 (15.43)
Fall DA	10.72 (2.61)	9.04 (3.24)
Spring Traditional Assessment		
Word Attack	10.92 (7.42)	14.84 (9.67)
WRAT Reading	21.76 (4.37)	24.81 (4.78)
Fluency	48.86 (31.34)	73.15 (33.43)
WIAT	12.72 (3.51)	16.53 (4.80)
CBM Sounds	54.84 (16.93)	
CBM Words	20.66 (22.16)	35.23 (21.52)
Spring DA	9.12 (2.51)	7.40 (3.36)

Table 6 and Table 7 display Pearson correlation coefficients between all measures used in the analysis. Correlations are displayed separately for kindergarten students (Table 6) and first grade students (Table 7).

For kindergarten students, the four reading outcome measures were statistically significantly correlated in the Spring. The predictor variables (Fall DA, CBM intercept, and CBM slope), by contrast, were inconsistently correlated. Fall DA and CBM intercept were statistically significantly correlated; however, CBM slope was not significantly correlated with either Fall DA or CBM intercept. With regard to the predictive correlations, CBM intercept was statistically significantly correlated with all four outcome measures; Fall DA was statistically significantly correlated with three outcome measures (WRAT Reading, Word Attack, and fluency); and, CBM slope was statistically significantly correlated with two outcome measures (Word Attack and fluency).

Table 6.

Kindergarten Correlation Matrix of Three Predictor Variables and Four Outcome Variables (N=25).

Measure	Fall DA	CBM Slope	CBM Intercept	WRAT Reading	Word Attack	Fluency	WIAT Spelling
Fall DA	1						
CBM Slope	-.352	1					
CBM Intercept	-.493*	.227	1				
WRAT Reading	-.624**	.280	.795**	1			
Word Attack	-.706**	.418*	.706**	.847**	1		
Fluency	-.585**	.415*	.921**	.873**	.763**	1	
WIAT Spelling	-.351	.069	.636**	.591**	.502*	.670**	1

Note: Slope and intercept based on CBM using letter sounds. (**) Correlation is significant at the 0.01 level. (*) Correlation is significant at the 0.05 level.

Table 7.

First Grade Correlation Matrix of Three Predictor Variables and Four Outcome Variables (n=95).

Measure	Fall DA	CBM Slope	CBM Intercept	WRAT Reading	Word Attack	Fluency	WIAT Spelling
Fall DA	1						
CBM Slope	-.353**	1					
CBM Intercept	-.625**	.569**	1				
WRAT Reading	-.745**	.612**	.744**	1			
Word Attack	-.765**	.495**	.673**	.846**	1		
Fluency	-.613**	.679**	.830**	.761**	.666**	1	
WIAT Spelling	-.636**	.554**	.704**	.704**	.699**	.718**	1

Note: Slope and intercept based on CBM using sight words. (**) Correlation is significant at the 0.01 level. (*) Correlation is significant at the 0.05 level.

For first grade students, the four reading outcome measures were very strongly correlated in the Spring. In contrast to the kindergarten data, the predictor variables (Fall DA, CBM intercept, and CBM slope) were also statistically significantly correlated. Correlations between predictor variables and outcome measures were more consistent in the first grade sample than the kindergarten sample. All three predictor variables were statistically significantly correlated with the four outcome measures.

Multiple Regression on Kindergarten Students' Spring Reading Performance

Kindergarten results from a multiple regression analysis are presented in Table 8. Results are discussed for each of the four dependent measures.

Table 8.

Multiple regression analysis using Fall DA, CBM intercept, and CBM slope to predict Spring reading performance for kindergarten students.

	Beta	t-value	Significance	Adjusted R^2 of Model
Kindergarten (N=25)				
WRAT Reading				
Constant	18.327	4.600	.000	
Fall DA	-.498	-2.089	.049	
CBM Intercept	.250	4.695	.000	
CBM Slope	.044	.236	.816	
				.662
Word Attack				
Constant	11.595	1.677	.108	
Fall DA	-1.196	-2.887	.009	
CBM Intercept	.305	3.299	.003	
CBM Slope	.416	1.273	.217	
				.647
Fluency				
Constant	-23.686	-1.441	.164	
Fall DA	-1.352	-1.373	.184	
CBM Intercept	2.302	10.463	.000	
CBM Slope	2.007	2.584	.017	
				.888
WIAT Spelling				
Constant	7.703	1.714	.101	
Fall DA	-.109	-.404	.691	
CBM Intercept	.194	3.221	.004	
CBM Slope	-.119	-.559	.582	
				.332

Fall DA and CBM intercept explained statistically significant variance in the prediction of letter knowledge and word reading as measured by the WRAT Reading subtest in the Spring. CBM slope was not a significant predictor. Overall, the model explained 66% of the variance in reading achievement. Fall DA and CBM intercept explained statistically significant variance in nonword reading as measured by the WRMT-R Word Attack subtest. CBM slope was not a significant predictor. The model explained 65% of the variance in reading achievement. CBM intercept and CBM slope explained statistically significant variance in fluency. Fall DA was not a significant predictor. The model explained 89% of the total variance in reading achievement. Only the CBM intercept explained significant variance in spelling as measured by the WIAT. Fall DA and CBM slope were not significant predictors. Overall, the model explained only 33% of the variance in spelling achievement.

Summary. The most consistent and significant predictor of kindergarten Spring reading performance was CBM intercept. CBM intercept explained statistically significant variance in all four reading measures. Fall DA was a statistically significant predictor for two reading variables (WRAT Reading and Word Attack), and CBM slope was a significant predictor for one dependent variable (fluency). The combination of the three independent variables predicted the most variance in fluency (89%), followed by WRAT Reading and Word Attack (66% and 65% respectively), and finally WIAT Spelling (33%).

Multiple Regression on First Grade Students' Spring Reading Performance

First grade results from a multiple regression analysis are presented in Table 9. Results are discussed for each of the four dependent measures.

Table 9.

Multiple regression analysis using Fall DA, CBM intercept, and CBM slope to predict Spring reading performance for first grade students.

	Beta	t-value	Significance	Adjusted R^2 of Model
First grade (N=95)				
WRAT Reading				
Constant	27.666	22.150	.000	
Fall DA	-.681	-6.660	.000	
CBM Intercept	.092	3.751	.000	
CBM Slope	.798	4.279	.000	
				.726
Word Attack				
Constant	25.609	9.050	.000	
Fall DA	-1.692	-7.302	.000	
CBM Intercept	.141	2.537	.013	
CBM Slope	.962	2.275	.025	
				.656
Fluency				
Constant	53.062	6.479	.000	
Fall DA	-1.605	-2.393	.019	
CBM Intercept	1.218	7.560	.000	
CBM Slope	6.073	4.964	.000	
				.759
WIAT Spelling				
Constant	17.409	11.226	.000	
Fall DA	-.479	-3.771	.000	
CBM Intercept	.116	3.816	.000	
CBM Slope	.650	2.806	.006	
				.580

Fall DA, CBM intercept, and CBM slope were all statistically significant predictors of WRAT Reading. The model explained 73% of the variance in Spring single word reading achievement. Fall DA, CBM intercept, and CBM slope were all statistically significant predictors of Word Attack. The model explained 66% of the total variance in Spring nonword reading achievement. Fall DA, CBM intercept, and CBM slope were all statistically significant predictors of fluency. The model explained 76% of the total variance in Spring oral reading fluency. And finally, Fall DA, CBM intercept, and CBM slope were all statistically significant predictors of WIAT Spelling. The model explained 58% of the total variance in Spring spelling achievement.

Summary. All three independent variables were consistent and significant predictors of the four dependent measures. The combination of the three independent variables predicted the most variance in fluency and WRAT Reading (76% and 73% respectively), followed by Word Attack (66%), and WIAT Spelling (58%).

Commonality Analysis

A commonality analysis was conducted to determine the unique contribution of each of the predictors and the common contribution among the predictors. This approach was developed by Mood (1969, 1971) and Mayeske et al. (1969) during the analysis of the Coleman Report (Coleman et al., 1966). The unique contribution of a predictor is the proportion of variance explained when it is entered last into the analysis. The common contribution is the proportion of variance explained by any one of the predictor variables. It is the shared variance among predictors.

Commonality analyses are particularly useful in predictive studies. A simple regression reports an R^2 statistic that represents the total amount of variance explained in the dependent variable by all the independent variables. Simple regression analyses cannot, however, partition the total variance (R^2) into portions of unique variance accounted for by each of the independent variables separately. Commonality analysis is particularly useful in studies of prediction because it helps researchers determine which variables may be eliminated without sacrificing overall predictability of the regression model. Variables that contribute the least amount of unique variance can sometimes be removed in a regression model without significantly reducing the amount of total variance explained.

For example, in the current study, three predictors were used: Fall DA, CBM intercept, and CBM slope. A simple regression analysis may determine that the model using three predictors explains 75% of the variance in the dependent variable. If a commonality analysis later reveals that Fall DA contributes an insignificant amount of unique variance, there would be no need to use both progress monitoring and Fall DA in the prediction of academic achievement. Progress monitoring alone (i.e., CBM intercept and CBM slope) could be used in the prediction of academic achievement, and the time and energy it takes to administer and score the DA would be saved. The commonality analysis, therefore, allows us to explore the added value of any particular predictor of interest. Table 10 reports the results of the commonality analysis for both kindergarten and first grade students.

Table 10.

Commonality analysis: Unique variance explained by Fall DA, CBM intercept, and CBM slope.

Measure	Age	
	Kindergarten	First Grade
	N=25	N=95
WRAT Reading		
Common	.054*	.234*
Fall DA unique	.052*	.129*
CBM Intercept unique	.323*	.039*
CBM Slope unique	-.014	.052*
Word Attack		
Common	.084*	.186*
Fall DA unique	.118*	.196*
CBM Intercept unique	.159*	.020*
CBM Slope unique	.010	.016*
Fluency		
Common	.100*	.214*
Fall DA unique	.004	.013*
CBM Intercept unique	.552*	.147*
CBM Slope unique	.029*	.062*
WIAT Spelling		
Common	.000	.180*
Fall DA unique	-.025	.060*
CBM Intercept unique	.285*	.062*
CBM Slope unique	-.021	.031*

Note: (*) significant amount of variance explained.

Commonality Analysis of Three Predictor Variables for Kindergarten Students

For the purposes of the following discussion, words that describe the relative sizes of common and unique variances (e.g., greater, bigger, more than, etc.) should not be understood as denoting a statistical comparison. These terms are only used to describe the relationships between the unique variances of the predictor variables. In addition, the amount of unique variance cannot legitimately be compared across outcome measures. For example, if Fall DA explains 5% unique variance and CBM intercept explains 32% unique variance in WRAT Reading, CBM intercept explains "more" unique variance than Fall DA. However, if CBM intercept explains 32% unique variance in WRAT reading and 16% unique variance in Word Attack, it cannot be stated that CBM intercept explains "more" unique variance in WRAT Reading than in Word Attack. The relative unique variance explained by each predictor can only be understood within the context of one outcome variable.

WRAT Reading. The regression model explained 66% of the total variance in kindergarten achievement on the WRAT Reading. The common variance explained was 5%. CBM intercept explained the greatest amount of unique variance (32%). The unique variance explained by Fall DA was 5%. CBM slope was not a statistically significant predictor of achievement on the WRAT Reading and did not explain any significant unique variance.

Word Attack. The regression model explained 65% of the total variance in kindergarten achievement on Word Attack. The common variance explained was 8%. CBM intercept explained the greatest amount of unique variance (16%). The unique variance explained by Fall DA was 12%. Again, CBM slope was not a statistically significant predictor of achievement on the Word Attack and did not explain any significant unique variance.

Fluency. The regression model explained 89% of the total variance in kindergarten achievement in fluency. The common variance explained was 10%. CBM intercept explained the greatest amount of unique variance (55%). CBM slope also explained a significant amount of unique variance (3%). Fall DA was not a statistically significant predictor of achievement in fluency and did not explain any significant unique variance.

WIAT Spelling. The regression model explained 33% of the total variance in kindergarten achievement on the WIAT Spelling. There was no common variance explained by the three predictors. CBM intercept was the only variable that explained a significant amount of unique variance (29%). Fall DA and CBM slope were not significant predictors of achievement on the WIAT Spelling and did not explain any significant unique variance.

Summary. For kindergarten students, CBM intercept was the most consistent predictor of reading achievement. It also consistently accounted for the greatest amount of unique variance across all dependent measures. Fall DA contributed unique variance to WRAT Reading and Word Attack and CBM slope contributed unique variance to fluency. The CBM intercept, therefore, was a useful tool in the prediction of a wide range of reading related skills (e.g., single word reading, nonword reading, fluency, and spelling), whereas Fall DA and CBM slope were useful tools in the prediction of specific skills (e.g., nonword reading for Fall DA and fluency for CBM slope). Fall DA contributed to the prediction of single word reading and nonword reading. CBM Slope contributed to reading fluency. These results may reflect the similarity of the Fall DA and the Spring Word Attack as well as the similarity of the CBM slope and Spring fluency. Fall DA is a nonword reading task and consequently predicted the greatest amount of unique variance on the nonword reading dependent variable Word Attack.

Similarly, CBM slope is based on a timed, rapid letter sound measure and consequently predicted the greatest amount of unique variance on the timed reading fluency dependent variable.

The common variance explained by the three predictor variables is often lower than the unique variance explained by one or two of the predictors. For example, the CBM intercept alone explains more unique variance than the variance than is common to all three predictors, indicating the relative importance of CBM intercept as a predictor of reading achievement. In another example (i.e., Word Attack), both Fall DA and CBM intercept explain unique variance higher than the common variance. Again, this is an indication that Fall DA and CBM intercept are both particularly important in the prediction of nonword reading.

Commonality Analysis of Three Predictor Variables for First Grade Students

WRAT Reading. The regression model explained 73% of the total variance in first grade achievement on the WRAT Reading. The common variance explained was 23%. Fall DA explained the greatest amount of unique variance (13%). CBM slope and CBM intercept also explained unique variance (5% and 4%, respectively).

Word Attack. The regression model explained 66% of the total variance in first grade achievement on Word Attack. The common variance explained was 19%. Fall DA explained the greatest amount of unique variance (20%). CBM slope and CBM intercept also explained unique variance (2% each).

Fluency. The regression model explained 76% of the total variance in first grade achievement in fluency. The amount of common variance explained was 21%. CBM intercept explained the greatest amount of unique variance (15%). CBM slope and Fall DA also explained unique variance (6% and 1%, respectively), though considerably less than intercept.

WIAT Spelling. The regression model explained 58% of the total variance in first grade achievement on the WIAT Spelling. The amount of common variance explained was 18%. Fall DA and CBM intercept explained the greatest amount of unique variance (6% each). The unique variance explained by CBM slope was 3%.

Summary. For all four dependent variables, the common variance explained by the three first grade predictor variables was higher than the common variance explained by the three kindergarten predictor variables. This greater commonality at first grade may be attributed to the higher correlations between predictor variables at first grade. Recall that first grade CBM slope was more strongly and consistently correlated with other predictive measures than kindergarten CBM slope (Table 6 and Table 7).

For first grade students, all three independent variables (Fall DA, CBM intercept, and CBM slope) were consistent predictors of reading achievement. The amount of common variance explained by any of the predictors, however, was consistently greater than any of their unique contributions (with the exception of Fall DA and Word Attack). A large amount of variance in first grade achievement, therefore, can be explained by any one of the three predictor variables. Of the three independent variables, Fall DA explained the greatest amount of unique variance in WRAT Reading and Word Attack, and both Fall DA and CBM intercept explained the same amount of variance on the WIAT Spelling. On the fluency measure, however, Fall DA explained the least amount of unique variance. These results again seem to demonstrate that the skill assessed in the Fall best predicts that same skill in the Spring.

CHAPTER V

DISCUSSION

The purpose of this study was to investigate the predictive validity of a DA reading measure. Specifically, the predictive validity of the DA reading measure was investigated in relation to the predictive validity of progress monitoring within an RTI framework. Two research questions guided this study. First, DA and progress monitoring were both explored independently to determine if they predicted reading achievement. Second, the amount of unique variance explained by DA and progress monitoring was explored to investigate their relative value in the prediction of reading achievement.

Mixed Results for DA

Results indicated that DA, initial performance (CBM intercept), and progress monitoring (CBM slope) are statistically significant predictors of Spring reading achievement for kindergarten and first grade students. These results vary in consistency across age groups and across dependent measures. For kindergarten students, simple regression analysis showed that CBM intercept was the most consistent predictor of Spring reading achievement, and it explained statistically significant variance in all four dependent measures. CBM slope explained significant variance only for fluency. Fall DA explained significant variance for WRAT Reading (word identification) and Word Attack (nonword reading). The commonality analysis revealed that CBM intercept explained the most unique variance in each of the four dependent measures. Fall DA contributed a significant, yet relatively small, amount of unique variance to WRAT Reading and a relatively large amount of unique variance to Word Attack. CBM slope contributed a significant, yet relatively small, amount of unique variance to fluency only. The common amount of variance explained by the three predictors was statistically significant for three of the dependent measures (WRAT Reading, Word Attack, and fluency). However, it was less than the unique variance accounted for by the CBM intercept. There was no statistically significant common variance explained on the WIAT Spelling measure. With the exception of WIAT Spelling, each dependent measure had some unique variance explained by Fall DA, CBM intercept, or CBM slope. The three screening measures (Fall DA, CBM intercept, and CBM slope), therefore, seem important predictors of reading achievement. But their predictive strength is dependent on the reading skill predicted (i.e., word reading, nonword reading, or fluency).

For first grade students, simple regression analysis showed that Fall DA, CBM intercept, and CBM slope each explained statistically significant variance in all four dependent measures. The commonality analysis for first grade students revealed a different pattern than that of the kindergarten students. For all but one dependent measure (Word Attack), the common variance among the three predictors was greater than the unique variance explained by any single predictor. In the case of Word Attack, the amount of common variance was still relatively large, but it was surpassed by the unique variance explained by Fall DA. The large amount of common variance suggests that Fall DA, CBM intercept, and CBM slope may be more closely related for first grade students than for kindergarten students, and that Fall DA, CBM intercept, and CBM slope may be more dependent upon the same skill set at first grade.

Factors Contributing to Mixed Results

Selection of Participants

Results varied greatly from kindergarten to first grade. Selection of participants may have affected these results both statistically and conceptually. Only relatively high-achieving kindergarten students were selected to participate. Thus, predictive analyses were conducted on a fairly homogenous group of kindergarten students. Statistically speaking, restriction of range may have limited the ability of Fall measures to predict performance on Spring measures for kindergarten students. That is, there may not have been enough variance among kindergarten students to explain. Conversely, a more even distribution of first grade participants was selected. Using a more heterogeneous group may have created more variance to explain at Spring and led to more consistent and significant results.

Conceptually speaking, it is possible that DA, CBM intercept, and CBM slope are more predictive of low-achieving students. Predictive effects may have been more consistent and statistically significant for first grade students because the effects were driven by the low-achieving students. If low-achieving kindergarten students were included in the study, effects may have been more consistent. Further analysis is necessary to explore this possibility.

Selection of Measures

The selection of measures may have also created inconsistency in the results between kindergarten and first grade students. CBM data were collected using a letter sounds measure for kindergarten students and a single word reading measure for first grade students. CBM slopes for kindergarten students tended to be more erratic, whereas CBM slopes for first grade students tended to be more linear. The erratic kindergarten slopes may have been due to the difficulty level of the CBM letter sounds measure. Most kindergarten students did not find the letter sounds measure challenging. Because the skill was mastered by most students, this timed measure became more like a measure of attention. It is possible that students with good attention scored consistently well, whereas students with poor attention scored inconsistently. CBM letter sounds may not be a sensitive progress monitoring measure for high-achieving kindergarten students.

Limitations of Study

No Treatment

The most noteworthy limitation of this study is the lack of a treatment in the conventional sense of the word. In this study, "treatment" was typical classroom intervention. Not only did we make no effort to strengthen classroom intervention, we did not observe typical classroom reading instruction. Students across classrooms in this study may have received significantly different instruction in terms of type of intervention and amount of intervention. Differences in teacher motivation and expertise may have also affected student achievement.

Sample Size and Selection

An important limitation is the sample size and selection of kindergarten students. Results based on a relatively homogenous group of 25 high-achieving kindergarten students should be interpreted very cautiously. Results based on a distribution of 95 first grade students are probably more reliable but here, too, the sample could have been larger.

Another problem was the timeline for selection of the sample. Due to time constraints, participants were selected quickly. Only those students who returned their consent forms before our screening procedure ended were allowed to participate. Students who returned consent forms quickly may be different from those who did not. If so, the external validity of these findings could be limited.

Timing of Assessments

In a typical school year, screening would be conducted within the first few weeks as an initial assessment of students' ability. In this study, CBM intercept, CBM slope, and Fall DA were measured from November to January, midway through the year. If these measures had been administered at the beginning of the school year, before students received any instruction, their predictive validity may have been different.

Contribution to Current Literature

Validity Explored

The purpose of this study was to explore the predictive validity of a DA reading measure. Results indicate that it is possible to examine predictive validity of standardized graduated prompt DA. Prediction of future achievement is important because it may identify the students who are at-risk for school failure and need more intensive intervention. Students enter school with different levels of background knowledge and different prognoses for immediate change. Whereas traditional assessment will reflect mostly a student's current knowledge, DA may be able to reflect both a student's current knowledge and a student's potential for change. Students with deficits in current knowledge but high potential for change may be less in need of immediate intensive intervention than students with deficits in current knowledge and low potential for change.

In addition to the poor reporting of reliability, studies of DA rarely report fidelity data regarding the administration and scoring of the DA measure. For this study, the project coordinator and I designed a fidelity protocol which measures the accuracy of each examiner's assessment. Before examiners conducted their school-based assessments, they were required to demonstrate 90% or above on this fidelity protocol. Even with this safeguard in place, however, problems in DA administration occurred. Monitoring audiotapes of the testing sessions and measuring inter-rater agreement were essential. By doing so, we were able to retest students or discard their data, depending on severity of the testing errors. Studies without fidelity data and inter-rater agreement should be interpreted most cautiously.

If DA is to become a viable method of assessment, it is essential that issues of fidelity, reliability, and validity be explored consistently.

Academic Relevance

One limitation of past DA research is its focus on general cognitive skills instead of academic skills. Campione and Brown conducted much of their graduated prompt DA research with cognitive skills (Campione et al., 1985a; Campione et al., 1985b), however, they believed that future DA research would be more useful as a tool to measure academic skills (Campione & Brown, 1987). If research can continue to demonstrate that DA has potential to inform educational decisions, such as placement, identification, or instructional planning, it may develop a stronger research base.

Possible Alternative to RTI

The reauthorization of IDEA (2004) allows for the first time the use of RTI to identify students with a specific learning disability. Most RTI models require anywhere from 10 to 30 weeks before a child can be considered a "nonresponder" and eligible for special education services. Using this model, children who will ultimately qualify for these services will not be receiving them during the 10 to 30 weeks of monitoring that RTI requires. Roughly half of the school year could pass without appropriately intensive intervention.

DA is a possible alternative method of identifying nonresponders. DA still assesses a student's "responsiveness," but it does so in a much shorter time frame (i.e., one testing session). It is possible that DA could be used as a screening measure within an RTI model. Students scoring very poorly on DA could be eligible for special education services faster. Instead of being monitored within a conventional "tier one" intervention, students scoring poorly on DA could immediately go on to a more intensive intervention over the course of 10 to 30 weeks while being monitored.

Considerations for Future Research

Contribution of Affective Factors

This analysis did not consider the contribution of affective factors, such as attention and motivation, which can influence learning. Because DA is often administered individually, examiners can manipulate attention and motivation more so than classroom teachers who must monitor many students. During the administration of DA in this study, examiners were allowed to redirect students' attention as necessary. Furthermore, students were motivated by the promise of a prize if they "worked hard" and "paid attention." These conditions do not closely resemble whole class instruction in schools. It may be interesting in future studies to investigate whether DA plus a measure of student attention predicts achievement better than DA alone.

Choosing Outcome Measures

In predictive validity studies, serious consideration should be given to the selection of outcome measures. The main question is, "What outcome are we trying to predict?" A related question is, "What are the skills most representative of that outcome?" In this study, we chose to investigate how well DA predicts individual children's reading achievement as measured by standardized tests. Performance on standardized tests, however, does not necessarily generalize to success or failure in the classroom. Perhaps curriculum-based outcome measures or teacher judgment of classroom achievement would be a more sensitive index of success in the classroom.

Regarding skills to be assessed, we chose in this study to investigate reading-related achievement by measuring single word reading, nonword reading, oral reading fluency, and spelling. Some may suggest that predicting nonword reading is less important than predicting single word reading and oral reading fluency. Selecting multiple measures using real words may be more appropriate in that case.

One final thought on the selection of measures concerns the relationship of the predictor variables to the outcome variables. If predictor A measures the same skill as the outcome measure and predictor B does not, it would naturally follow that predictor A is the stronger of the two. Selecting varied outcome measures, therefore, may be important to keep the magnitude of the results in perspective. For example, if only WRAT Reading and Word Attack were used as outcome measures in this study, I may have concluded that DA was a stronger predictor of Spring reading achievement. If only fluency was used as an outcome measure, I may have concluded that CBM intercept and CBM slope were stronger. Only by using multiple measures, I found that DA, CBM intercept, and CBM slope were strong predictors of Spring reading achievement but they predicted different reading skills. DA was a stronger predictor of nonword reading and single word reading. CBM intercept and CBM slope were stronger predictors of oral reading fluency.

The Link between Assessment and Intervention

DA has been described as a more educationally valid assessment measure because of its ability to inform instruction (Campione & Brown, 1987; Lidz et al., 1997). This contention has been studied extensively by Feuerstein (Feuerstein et al., 1979a, 1979b; Rand, Tannenbaum, & Feuerstein, 1979), though not empirically. Future researchers may want to consider designing specific “matched” interventions based on a student’s pretest DA performance. Then, by randomly assigning students to either a “matched” or "mismatched" treatment group, we can begin to investigate whether DA can live up to its promise.

To illustrate a possible example from this study, consider the CVCE subtest of the DA. If we found that certain children failed items because they did not know their letter sounds, lower level instruction on letter identification and phonological awareness may be appropriate. If we found that certain children failed items because they had difficulty understanding the long "o" rule, instruction on learning and generalizing reading rules may be appropriate. If we found that certain children failed because they had difficulty attending to the task, instruction that includes positive behavioral reinforcement for attention may be appropriate. If we found that certain children failed because they had difficulty understanding the relevance of reading nonwords, meaningful instruction that focuses on rule learning using real words may be appropriate. There are many possibilities. Future studies must carefully attend to the supposed reasons for student failure, design interventions based on different types of failure, and test these interventions empirically.

Appendix A

Dynamic Assessment
Fidelity Checklist Static Measures

Tester: ____________________________ Observer: ______________________________

Time and Date: __________________________

Reliablity is defined by 90% or above

General Testing Behaviors

+	—	NA		**General Testing Behaviors**
			1	Tester positions clipboard appropriately. (i.e. student unable to see scoring)
			2	Tester is positive and smiles a great deal.
			3	Tester praises for effort and not for correct responses.
			4	Tests always face the student.
			5	Tests are covered appropriately.
			6	Test administered in the correct order. (i.e. random order)
			7	Tester records from beginning making sure to record student's first/last name
A = Total (+)	B= Total (-)	C = Total (+) and (-)	A/C * 100 = % accuracy	Comments:
General Testing Behavior Fidelity				

Letter Sounds

+	—	NA		Letter Sounds
			1	Tester gives initial directions verbatim.
			2	Tester uses the appropriate correction procedure.
			3	Tester gives middle directions verbatim.
			4	Tester moves student along after 3 seconds by pointing to the next sound.
			5	Tester praises student for good effort.
A = Total (+)	B= Total (-)	C = Total (+) and (-)	A/C * 100 = % accuracy	Comments:
Letter Sound Fidelity				

Decoding Inventory

+	—	NA		**Decoding Inventory**
			1	Tester reads directions verbatim.
			2	Tester corrects student when appropriate.
			3	Tester encouraged student twice to sound the words out if letter names or
				if real words were stated. (Warning can be given 2 times)
			4	Tester moves student along after 5 seconds by pointing to the next sound.
			5	Tester praises student for good effort.
A = Total (+)	B= Total (-)	C = Total (+) and (-)	A/C * 100 = % accuracy	Comments:
Decoding Inventory Fidelity				

Rapid Letter Name

+	—	NA		**Rapid Letter Name (RLN)**
			1	Tester gives directions verbatim.
			2	If student does not respond the tester gives letter after 3 seconds.
			3	If student gives incorrect response the tester does not correct.
			4	Test is administered for 60 seconds.
			a	If student begins by stating 3 letter sounds in a row at the beginning, tester gives warning and starts timer and test over again.
			5	If student states 3 consecutive letter sounds anywhere in the test, other than
				beginning, tester gives warning but does not restart timer.
			6	Tester moves student along after 3 seconds by pointing to the next sound.
			7	Tester praises student for good effort.
A = Total (+)	B= Total (-)	C = Total (+) and (-)	A/C * 100 = % accuracy	Comments:
Rapid Letter Name Fidelity				

Segmenting

+	—	NA		Segmenting
			1	Tester gives directions verbatim.
			2	Tester gives all three sample items.
			3	Tester uses 3 fingers to indicate sounds
			4	Tester starts the timer after she says, "say the sounds in dog."
			5	Tester says, "Say the sounds in…" before each item.
			6	Tester corrects when appropriate.
			a	dog
			b	fine
			c	she
			d	grew
			e	red
			f	sat
			g	lay
			h	zoo
			i	job
			j	ice
			k	top
			l	do
			m	keep
			n	no
			o	wave
			7	Test is administered for 60 seconds.
			8	Tester moves student along after 3 seconds by moving to the next item.
			9	Tester praises student for good effort.
A = Total (+)	B= Total (-)	C = Total (+) and (-)	A/C * 100 = % accuracy	Comments:
Segmenting Fidelity				

WRMT - R Word Attack

+	—	NA		**WRMT - R Word Attack**
			1	Tester gives directions verbatim.
			2	If necessary, tester corrects sample items.
			a	tat
			b	op
			3	Tester moves student along after 5 seconds by pointing to the next word.
			4	Tester administers until 6 consecutive wrong answers are given.
			5	Tester praises effort.
A = Total (+)	B= Total (-)	C = Total (+) and (-)	A/C * 100 = % accuracy	Comments:
WRMT - R Word Attack Fidelity				

WRAT Reading

+	—	NA		**WRAT Reading**
			1	Tester gives letter directions verbatim.
			2	Tester corrects first error.
			3	Tester moves along after 10 seconds by pointing to the next letter.
			4	Tester gives word directions verbatim.
			5	Tester moves along after 10 seconds by pointing to the next word.
			6	Tester administers until 10 consecutive incorrect responses.
			7	Tester praises effort.
A = Total (+)	B= Total (-)	C = Total (+) and (-)	A/C * 100 = % accuracy	Comments:
WRAT Reading Fidelity				

WRAT Arithmetic

+	—	NA		**WRAT Arithmetic**
			1	Tester gives directions for the oral section verbatim.
			2	Tester gives oral item directions verbatim.
			a	3 ducks
			b	5 boxes
			c	15 dots
			d	3
			e	5
			f	6
			g	17
			h	41
			i	3 fingers
			j	8 fingers
			k	9
			l	42
			m	2 pennies
			n	7 apples
			o	6 marbles
			3	Tester gives direction for the written section verbatim.
			4	Tester sets the timer for 10 minutes.
			5	If student that he/she is finished tester says, "Are there any of these problems
				that you think that you can do?"
			6	If the student works the entire 10 mintues the tester says, "Stop! Put your
				pencil down."
			7	Tester praises effort
Total (+)	B= Total (-)	(+) and (-)	= % accuracy	Comments:
WRAT Arithmetic Fidelity				

CBM

+	—	NA		**CBM**
			1	Tester gives practice item directions verbatim.
			2	Tester administers practice items.
			3	Tester administers scored test directions.
			4	List 1
			a	Tester starts timer after directions.
			b	Tester prompts student after 2 seconds by saying, "Go on."
			c	Tester ends the test after 60 seconds by saying, "Stop."
			5	List 2
			a	Tester starts timer after directions.
			b	Tester prompts student after 2 seconds by saying, "Go on."
			c	Tester ends the test after 60 seconds by saying, "Stop."
			6	Tester praises effort.

A = Total (+)	B= Total (-)	C = Total (+) and (-)	A/C * 100 = % accuracy
CBM Fidelity			

Comments:

Appendix B

Dynamic Assessment
Fidelity Checklist Dynamic Assessment

Tester: ____________________________ Observer: ______________________________

Time and Date: __________________________

Reliablity is defined by 90% or above

Dynamic Assessment

+	—	NA		**CVC**
				Level 1
			1	Tester delivers instructions verbatim.
			2	Tester prompts student to read nonsense words.
				Level 2
			1	Tester delivers instructions verbatim.
			2	Tester sorts words appropriately and helps the child to sort the words.
			3	Tester prompts student to read nonsense words.
				Level 3
			1	Tester delivers instructions verbatim.
			2	Tester sorts words appropriately and helps the child to sort the words.
			3	Tester prompts student to read nonsense words. (Tester must say
				"remember what you've just learned to help you read these nonsense words")
				Level 4
			1	Tester delivers instructions verbatim.
			2	Tester prompts student to read nonsense words.
				Level 5
			1	Tester delivers instructions verbatim.
			2	Tester prompts student to be the teacher.
			3	Tester plays "guess my word."
			4	Tester prompts student to read nonsense words. (Tester must say
				"remember what you've just learned.")
Total (+)	B= Total (-)	(+) and (-)	= % accuracy	Comments:
CVC Fidelity				

Dynamic Assessment

+	−	NA		**CVCE**
				Level 1
			1	Tester delivers instructions verbatim.
			2	Tester prompts student to read nonsense words.
				Level 2
			1	Tester delivers "number of sounds in a word" instructions verbatim.
			2	Tester delivers "is the middle vowel sound the same or different" instructions.
			3	Tester sorts words under "yes and no" cards.
			4	Tester prompts student to read nonsense words.
				Level 3
			1	Tester delivers instructions verbatim.
			2	Tester sorts words under "long o, short o" cards. (asking the child for help.)
			3	Tester prompts student to read nonsense words. (Tester must say
				"remember what you've just learned and please try and read these nonsense
				words")
				Level 4
			1	Tester delivers instructions verbatim.
			2	Tester prompts student to read nonsense words. (Tester must say, "think
				about what we've just talked about and read these nonsense words.")
				Level 5
			1	Tester delivers instructions verbatim.
			2	Tester uses the word cards to ask the student if "o" says its own name.
			3	Tester prompts student to read nonsense words.

Total (+)	B= Total (-)	(+) and (-)	= % accuracy
CVCE Fidelity			

Comments:

Dynamic Assessment

+	−	NA		**Doubling**
				Pre-teaching
			1	Tester conducts "ing" teaching until student displays 100% accuracy.
				Level 1
			1	Tester delivers instructions verbatim.
			2	Tester prompts student to read nonsense words.
				Level 2
			1	Tester delivers the "number of sounds" instructions verbatim.
			2	Tester delivers the "is the vowel sound the same or different" instructions
				verbatim.
			3	Tester delivers the "long and short vowel terminology" instructions verbatim.
			4	Tester prompts student to read the nonsense words. (tester must say,
				"Remember to look carefully at each word as you read these nonsense words.
				Level 3
			1	Tester delivers instructions verbatim.
			2	Tester prompts student to read nonsense words. (Tester must say, "think
				about the words I just read and try real hard to read these nonsense words.")
				Level 4
			1	Tester delivers initial instructions verbatim.
			2	Tester delivers "one t" practice instructions verbatim.
			3	Tester delivers "two d" practice instructions verbatim.
			4	Tester delivers "one d / two d" practice instructions verbatim.
			5	Tester prompts student to read nonsense words. (Tester must say, "Think
				about what we just talked about. Please look carefully at these words and do
				your best to read them."
				Level 5
			1	Tester delivers initial instructions verbatim.
			2	Tester delivers "one t" practice instructions verbatim.
			3	Tester delivers "two d" practice instructions verbatim.
			4	Tester prompts student to read nonsense words. (Tester must say, "You are
				really working hard for me. Please try your best to read these nonsense
				words.

Total (+)	B= Total (-)	(+) and (-)	= % accuracy	Comments
Doubling Fidelity				

REFERENCES

Anastasi, A. (1988). *Psychological testing* (6th ed.). New York: Macmillan.

* Babad, E.Y., & Budoff, M. (1974). Sensitivity and validity of learning-potential measurement in three levels of ability. *Journal of Educational Psychology, 66,* 439-447.

* Bain, B.A., & Olswang, L.B. (1995). Examining readiness for learning two-word utterances by children with specific expressive language impairment: Dynamic assessment validation. *American Journal of Speech Language Pathology, 4*, 81-91.

Bransford, J.C., Delclos, J.R., Vye, N.J., Burns, M., & Hasselbring, T.S. (1987). State of the art and future directions. In C.S. Lidz (Ed.), *Dynamic assessment: An interactional approach to evaluating learning potential* (pp. 479-496). New York: Guilford Press.

* Bryant, N.R. (1982). *Preschool children's learning and transfer of matrices problems: A study of proximal development*. Unpublished master's thesis, University of Illinois.

* Bryant, N.R., Brown, A.L., & Campione, J.C. (1983, April). *Preschool children's learning and transfer of matrices problems: Potential for improvement*. Paper presented at the meeting of the Society for Research in Child Development, Detroit.

Budoff, M. (1967). Learning potential among institutionalized young adult retardates. *American Journal of Mental Deficiency, 72,* 404-411.

Budoff, M. (1987a). Measures for assessing learning potential. In C.S. Lidz (Ed.), *Dynamic Testing* (pp.173-195). New York: Guilford Press.

Budoff, M. (1987b). The validity of learning potential. In C.S. Lidz (Ed.), *Dynamic Testing* (pp. 52-81). New York: Guilford Press.

Budoff, M., Corman, L., & Gimon, A. (1976). An educational test of learning potential testing with Spanish-speaking youth. *Interamerican Journal of Psychology, 10*, 13-24.

* Budoff, M., Gimon, A., & Corman, L. (1974). Learning potential measurement with Spanish-speaking youth as an alternative to IQ tests: A first report. *Interamerican Journal of Psychology, 8,* 233-246.

Budoff, M., & Friedman, M. (1964). “Learning potential” as a testing approach to the adolescent mentally retarded. *Journal of Consulting Psychology, 28,* 434-439.

* Budoff, M., Meskin, J., & Harrison, R.H. (1971). Educational test of the learning-potential hypothesis. *American Journal of Mental Deficiency, 76*, 159-169.

* Byrne, B., Fielding-Barnsley, R., & Ashley, L. (2000). Effects of preschool phoneme identity training after six years: Outcome level distinguished from rate of response. *Journal of Educational Psychology, 92,* 659-667.

Campione, J.C., & Brown, A.L. (1987). Linking dynamic assessment with school achievement. In C.S. Lidz (Ed.) *Dynamic Testing* (pp. 82-115). New York: Guilford Press.

Campione, J.C., & Brown, A.L., & Bryant, N. (1985a). Individual differences in learning and memory. In R.J. Sternberg (Ed.), *Human abilities: An information-processing approach* (pp. 103-126). New York: Freeman.

Campione, J.C., Brown, A.L., Ferrara, R.A., Jones, R.S., and Steinberg, E. (1985b). Breakdowns in flexible use of information: Intelligence-related differences in transfer following equivalent learning performance. *Intelligence, 9*, 297-315.

Carlson, J.S., & Wiedel, K.H. (1978). Use of testing-the-limits procedures in the testing of intellectual capabilities in children with learning difficulties. *American Journal of Mental Deficiency, 11*, 559-564.

Carlson, J.S., & Wiedel, K.H. (1979). Toward a differential testing approach: Testing-the-limits employing the Raven matrices. *Intelligence, 3,* 323-344.

Carlson, J.S., & Wiedl, K.H. (1980). Applications of a dynamic testing approach in intelligence assessment: Empirical results and theoretical formulations. *Zeitschrift fur Differentielle und Diagnostische Psychologie, 1*, 303-318.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences (Rev Ed)*. New York: Academic Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd Ed).* Hillsdale, NJ: Erlbaum.

Coleman, J.S., Campbell, E.Q., Hobson, C.J., McPartland, J., Mood, A.M., Weinfeld, F.D., & York, R.L. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Government Printing Office.

* Day, J.D., Engelhardt, J.L., Maxwell, S.E., & Bolig, E.E. (1997). Comparison of static and dynamic assessment procedures and their relation to independent performance. *Journal of Educational Psychology, 89*, 358-368.

Dillon, R. (1979). Improving validity by testing for competence: Refinement of a paradigm and its application to the hearing impaired. *Educational and psychological measurement, 39,* 363-371.

* Ferrara, R.A. (1987). *Learning mathematics in the zone of proximal development: The importance of flexible use of knowledge*. Unpublished doctoral dissertation, University of Illinois.

Ferrara, R.A., Brown, A.L., & Campione, J.C. (1986). Children's learning and transfer of inductive reasoning rules: Studies in proximal development. *Child Development, 57,* 1087-1099.

Feuerstein, R., Rand, Y., & Hoffman, M.B. (1979a). *The dynamic assessment of retarded performers: The Learning Potential Assessment Device*. Baltimore: University Park Press.

Feuerstein, R., Rand, Y., Hoffman, M., Hoffman, M., & Miller, R. (1979b). Cognitive modifiability in retarded adolescents: Effects of instrumental enrichment. *American Journal of Mental Deficiency, 83*, 539-550.

Feuerstein, R., Rand, J., & Rynders, J.E. (1988). *Don't accept me as I am: Helping "retarded" people to excel*. New York: Plenum Press.

Flammer, A., & Schmid, H. (1982). Lerntests: Konzept, Realisierungen, Bewährung. Eine Übersicht. *Schweizerische Zeitschrift für Psychologie und ihre Anwendungen, 33,* 14-32.

Fuchs, D., Fuchs, L.S., & Compton, D.L. (2006). Dynamic Assessment of Early Reading, Revised. Vanderbilt University.

Ginzburg, M.P. (1981). O vozmozhnoi interpretatsii poniatia zony blitzhaishego razvitia [On a possible interpretation of the concept of the zone of proximal development]. In D.B. El'konin & A.L. Venger (Eds.), *Diagnostika uchebnoi diatel'nosti I intellectual'nogo razvitia detei* (pp. 145-155). Moscow: Academia Pedagogicheskikh Nauk SSSR.

Goncharova, E.L. (1990). Nekotorye voprosy vyshego obrazovania vzrnslykh slepoglukhikh [On higher education for the deaf-blind]. In V.N. Chulkov, V.I. Lubovsky, & E.N. Martsinovskaia (Eds.), *Differentsirovannyi podkhod pri obuchenii I vospitanii slepoglukhikh detei* (pp. 56-70). Moscow: Academia Pedagogicheskikh Nauk SSSR.

Grigorenko, E.L., & Sternberg, R.J. (1998). Dynamic testing. *Psychological Bulletin, 124,* 75-111.

Guthke, J. (1977). *Zur Diagnostik der intellekturllen Lernahigkeit* [Assessment of intellectual learning potential]. Berlin: VEB Deutcher Verlag der Wissenschafen.

Hamers, J.H.M., Hessels, M.G.P., & Van Luit, J.E.H. (1991). Leertest voor etnische minderheden. *Test en handleiding.* Lisse: Swets & Zeitlinger.

Hamers, J.H.M., & Ruijssenaars, A.J.J.M. (1984). *Leergeschiktheid en leertests*. Lisse: Swets & Zeitlinger.

Haywood, H.C., Brown, A.L., & Wingenfeld, S. (1990). Dynamic approaches to psychoeducational assessment. *School Psychology Review, 19*, 411-422.

* Hessels, M.G.P., & Hamers, J.H.M. (1993). The learning potential test for ethnic minorities. In J.H.M. Hamers, K. Sijtsma, & A.J.J.M. Ruijssenaars (Eds.), *Learning Potential Assessment: Theoretical, Methodological, and Practical Issues* (pp.285-311), Lisse: Swets & Zeitlinger B.V.

Kaufman, A. (1975). Factor analysis of the WISC-R at 11 age levels and between 6½ and 16½ years. *Journal of Consulting and Clinical Psychology, 43,* 135-147.

Kratochwill, T.R., & Severson, R.A. (1977). Process assessment: An examination of reinforcer effectiveness and predictive validity. *Journal of School Psychology, 5,* 293-300.

Lidz, C.S. (Ed.) (1987). *Dynamic testing: An integrated approach in evaluating learning potential*. New York: Guilford Press.

* Lidz, C.S., Jepsen, R.H., & Miller, M.B. (1997). Relationships between cognitive process and academic achievement: Application of a group dynamic assessment procedure with multiply handicapped adolescents. *Educational and Child Psychology, 14,* 56-67.

Lipsey, M.W., & Wilson, D.B. (2000). *Practical meta-analysis*. Thousand Oaks, CA: Sage Publications.

Luther, M. & Wyatt, F. (1996). A comparison of Feuerstein's method of (LPAD) assessment with conventional IQ testing on disadvantaged North York high school students. In M. Luther, E. Cole, & P. Gamlin (Eds.), *Dynamic assessment for instruction: From theory to application* (pp.168-181). North York, ON, England: Captus Press, Inc.

Mayeske, G.W., Cohen, W.M., Wisler, C.E., Okada, T., Beaton, A.E., Proshek, J.M., Weinfeld, F.D., & Tabler, K.A. (1969). *A study of our nation's schools*. Washington, DC: U.S. Department of Health, Education, and Welfare, Office of Education.

Meijer, J. (1993). Learning potential, personality characteristics, and test performance. In J.H.M. Hamers, K. Sijtsma, & A.J.J.M. Ruijssenaars (Eds.), *Learning Potential Assessment: Theoretical, Methodological, and Practical Issues* (pp.341-362), Lisse: Swets & Zeitlinger B.V.

Mood, A.M. (1969). Macro-analysis of the American educational system. *Operations Research, 17*, 770-784.

Mood, A.M. (1971). Partitioning variance in multiple regression analyses as a tool for developing learning models. *American Educational Research Journal, 8*, 191-202.

Muttart, K. (1984). Assessment of effects of instrumental enrichment cognitive training. *Special Education in Canada, 58*, 106-108.

* Olswang, L.B., & Bain, B.A. (1996). Assessment information for predicting upcoming change in language production. *Journal of Speech and Hearing Research, 39*, 414-423.

* Peña, E., Quinn, R., & Iglesias, A. (1992). The application of dynamic methods to language assessment: A nonbiased procedure. *The Journal of Special Education, 26*, 269-280.

Rand, Y., Tannenbaum, A.J., & Feuerstein, R. (1979). Effects of instrumental enrichment on the psychoeducational development of low-functioning adolescents. *Journal of Educational Psychology, 71*, 751-763.

* Resing, W.C.M. (1993). Measuring inductive reasoning skills: The construction of a learning potential test. In J.H.M. Hamers, K. Sijtsma, & A.J.J.M. Ruijssenaars (Eds.), *Learning Potential Assessment: Theoretical, Methodological, and Practical Issues* (pp.219-242), Lisse: Swets & Zeitlinger B.V.

Ruijssenaars, A.J.J.M., Castelijns, J.H.M., & Hamers, J.H.M. (1993). The validity of learning potential tests. In J.H.M. Hamers, K. Sijtsma, & A.J.J.M. Ruijssenaars (Eds.), *Learning Potential Assessment: Theoretical, Methodological, and Practical Issues* (pp.69-82), Lisse: Swets & Zeitlinger B.V.

* Rutland, A., & Campbell, R. (1995). The validity of dynamic assessment methods for children with learning difficulties and nondisabled children. *Journal of Cognitive Education, 5*, 81-94.

* Samuels, M.T., Killip, S.M., MacKenzie, H., & Fagan, J. (1992). Evaluating preschool programs: The role of dynamic assessment. In H.C. Haywood and D. Tzuriel (Eds.) *Interactive Assessment*, (pp.251-271). New York: Springer-Verlag.

Sattler, J. (1988). *Assessment of children* (3rd ed.). San Diego: CA: Author.

Savell, J.M., Twohig, P.T., Rachford, D.L. (1986). Empirical status of Feuerstein's 'Instrumental Enrichment' (FIE) technique as a method of teaching thinking skills. *Review of Educational Research, 56*, 381-409.

* Sewell, T.E. (1979). Intelligence and learning tasks as predictors of scholastic achievement in black and white first-grade children. *Journal of School Psychology, 17*, 325-332.

* Sewell, T.E., & Severson, R.A. (1974). Learning ability and intelligence as cognitive predictors of achievement in first-grade black children. *Journal of Educational Psychology, 66*, 948-955.

Shocket, I. (1992). A dynamic assessment for undergraduate admission: The inverse relationship between modifiability and predictability. In H.C. Haywood and D. Tzuriel (Eds.) *Interactive Assessment*, (pp.332-355). New York: Springer-Verlag.

* Spector, J.E. (1992). Predicting progress in beginning reading: Dynamic assessment of phonemic awareness. *Journal of Educational Psychology, 84,* 353-363.

* Speece, D.L., Cooper, D.H., & Kibler, J.M. (1990). Dynamic assessment, individual differences, and academic achievement. *Learning and Individual Differences, 2*, 113-127.

* Swanson, H.L. (1994). The role of working memory and dynamic assessment in the classification of children with learning disabilities. *Learning Disabilities Research & Practice, 9*, 190-202.

* Swanson, H.L. (1995). Effects of dynamic testing on the classification of learning disabilities: The predictive and discriminant validity of the Swanson-Cognitive Processing Test (S-CPT). *Journal of Psychoeducational Assessment*, *13*, 204-229.

Tellegen, P.J., & Laros, J.A. (1993). The Snijders-Oomen nonverbal intelligence tests: General intelligence tests or tests for learning potential? In J.H.M. Hamers, K. Sijtsma, & A.J.J.M. Ruijssenaars (Eds.), *Learning Potential Assessment: Theoretical, Methodological, and Practical Issues* (pp.267-283), Lisse: Swets & Zeitlinger B.V.

Tharp, R.G., & Gallimore, R. (1988). *Rousing minds to life: Teaching, learning, and schooling in social context*. Cambridge, England: Cambridge University Press.

* Tissink, J., Hamers, J.H.M., & VanLuit, J.E.H. (1993). Learning potential tests with domain-general and domain-specific tasks. In J.H.M. Hamers, K. Sijtsma, & A.J.J.M. Ruijssenaars (Eds.), *Learning Potential Assessment: Theoretical, Methodological, and Practical Issues* (pp.243-266), Lisse: Swets & Zeitlinger B.V.

Utley, C.A., Haywood, H.C., & Masters, J.C. (1992). Policy implications of psychological assessment of minority children. In H.C. Haywood & D. Tzuriel (Eds.), *Interactive assessment* (pp. 445-469). New York: Springer-Verlag.

Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale-Revised*. San Antonio: TX: The Psychological Corporation.

Wiedl, K.H., & Herrig, D. (1978). *Okologische Validitat und Schulerfolgsprognose im Lern- und Intelligenztest: Eine exemplarische Studie, 24*, 175-186.